HHS Public Access
Author manuscript
Med Phys. Author manuscript; available in PMC 2018 February 01.

Published in final edited form as:
Med Phys. 2017 February ; 44(2): 547–557. doi:10.1002/mp.12045.

Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks

Bulat Ibragimov[1] and **Lei Xing**
Department of Radiation Oncology, Stanford University School of Medicine, Stanford, California 94305

Abstract

Purpose—Accurate segmentation of organs-at-risks (OARs) is the key step for efficient planning of radiation therapy for head and neck (HaN) cancer treatment. In the work, we proposed the first deep learning-based algorithm, for segmentation of OARs in HaN CT images, and compared its performance against state-of-the-art automated segmentation algorithms, commercial software and inter-observer variability.

Methods—Convolutional neural networks (CNNs) – a concept from the field of deep learning – were used to study consistent intensity patterns of OARs from training CT images and to segment the OAR in a previously unseen test CT image. For CNN training, we extracted a representative number of positive intensity patches around voxels that belong to the OAR of interest in training CT images, and negative intensity patches around voxels that belong to the surrounding structures. These patches then passed through a sequence of CNN layers that captured local image features such as corners, end-points and edges, and combined them into more complex high-order features that can efficiently describe the OAR. The trained network was applied to classify voxels in a region of interest in the test image where the corresponding OAR is expected to be located. We then smoothed the obtained classification results by using Markov random fields algorithm. We finally extracted the largest connected component of the smoothed voxels classified as the OAR by CNN, performed dilate-erode operations to remov cavities of the component, which resulted in segmentation of the OAR in the test image.

Results—The performance of CNNs was validated on segmentation of spinal cord, mandible, parotid glands, submandibular glands, larynx, pharynx, eye globes, optic nerves and optic chiasm using 50 CT images. The obtained segmentation results varied from 37.4% Dice coefficient (DSC) for chiasm to 89.5% DSC for mandible. We also analyzed the performance of state-of-the-art algorithms and commercial software reported in the literature, and observed that CNNs demonstrate similar or superior performance on segmentation of spinal cord, mandible, parotid glands, larynx, pharynx, eye globes and optic nerves, but inferior performance on segmentation of submandibular glands and optic chiasm.

Conclusion—We concluded that convolution neural networks can accurately segment most of OARs using a representative database of 50 HaN CT images. At the same time, inclusion of

[1] Corresponding author: bulat@stanford.edu.

The authors have no relevant conflicts of interest to disclose.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

additional information, e.g. MR images, may be beneficial for some OARs with poorly-visible boundaries.

Keywords

Deep learning; convolutional neural networks; head and neck; segmentation; radiotherapy

1. Introduction

Head and neck (HaN) cancer including oral cavity, salivary glands, paranasal sinuses and nasal cavity, nasopharynx, oropharynx, hypopharynx, and larynx cancers is among the most prevalent cancer types worldwide[1]. In the last decades, high precision radiation therapy such as intensity-modulated radiation therapy (IMRT), volumetric modulated radiation therapy (VMAT), and proton therapy have been widely used for HaN cancer treatment due to their ability for highly conformal dose delivery. To minimize post-treatment complications, organs-at-risks (OARs), such as brainstem, spinal cord, mandible, larynx, pharynx, parotid and submandibular glands, and, in the case of nasopharyngeal cancer, eyes, optic nerves and chiasm, must be accurately delineated. The complexity of OARs morphology and imperfection of imaging devices make manual delineation prone to errors and time consuming[2]. There is therefore a great demand for more accurate OARs delineation and for considerable reduction of the amount of manual labor in HaN treatment planning[3].

Computed tomography (CT)-based treatment planning remains to be the mainstay in current clinical practice for its high acquisition speed, high spatial accuracy and resolution, and the ability of providing relative electron density information. However, CT images have low contrast of soft tissues and are usually corrupted by metallic artifacts, which limits applicability of intensity-based segmentation algorithms. As the general appearance of OARs and surrounding tissues remains similar among CT images of different patients, there is a certain confidence that a previously analyzed training images can be used to segment a new test image. The most straightforward approach is to align, i.e. register, training images, so-called atlases, to the test image and correspondingly propagate the training OARs annotations. Such atlasing has been receiving considerable attention in the literature on HaN segmentation[3]. Non-rigid registration of three-dimensional (3D) images is a relatively slow procedure, therefore early HaN segmentation methods registered a single atlas with annotation to the test image [4–6]. This atlas was usually defined using an arbitrary training image, which negatively affected registration performance if the selected image was not representative[4, 6, 7]. Such bias can be reduced when the atlas is defined using an artificial image with standard anatomy[8, 9] or the average image constructed by merging several training images[5, 10, 11]. It is also possible to choose a new atlas image for every segmentation according to its similarity to the test image[12]. Single atlas however cannot guarantee accurate segmentation results in the presence of pathologies and tumors when the morphology of OARs is not similar enough between the atlas and test image. Multi-atlas segmentation is less sensitive to inter-patient anatomy variability and produces more accurate segmentation results if a high number of atlas images is used[13–15]. The registrations for all atlases are usually combined with simultaneous truth and performance level estimation (STAPLE)[16], similarity and truth estimation (STEPS)[16] and joint weighted

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

voting[17] techniques. When validated on HaN muscles, lymph nodes, spinal cord and brainstem, multi-atlas segmentation was shown to be slower but more accurate than single atlas segmenation[12]. Apart from atlas selection, the results strongly depend on the parameters of the applied registration algorithm. In the literature, registration of two HaN images was based on simple intensity-[18], mutual information-[4, 15, 19–21] and Gaussian local correlation coefficient-based[4] similarly measures, and optical flow[8], consistent volumes[22], scale-invariant features[23], locally affine block matching[5, 10, 16], demons[24], fluid dynamics[25] radial basis function[26] and B-splines[11, 13, 15, 20] for controlling registration transformations. It was also concluded that HaN registration convergence improves considerably when anchor landmarks located at anatomically relevant regions such as vertebrae and mandible corners are used[20, 27, 28]. Landmark-assisted atlasing demonstrated better performance in comparison to standard atlasing on segmentation of lymph nodes[20], brainstem[27] and individual tongue muscles[28].

Alternative HaN segmentation approaches applied fast marching[29], mass springs for lymph nodes modeling[30] parametrical shapes for eye globes and optical nerve modeling[31], active contours[32], deformable meshes[24, 33], principal component-based shapes[11, 19], graph cut[13, 17] and superpixels[15]. In general, shape-based approaches demonstrated exceptional performance on segmentation of heart[34] and spine[35–37], and annotation of X-ray head images[38, 39]. Researchers also studied performance of machine learning algorithms such as k-nearest neighbors and support vector machines trained on intensity, gradient, texture contrast, texture homogeneity, texture energy, cluster tendency, Gabor and Sobel features[21, 24]. The above mentioned algorithms covered the complete set of OARs in the HaN region including brainstem[4, 7–12, 14, 16–18, 24–27, 40, 41], cerebellum[9, 13, 14, 17, 25, 41], spinal cord[4, 13, 16, 18, 40, 41], mandible[4, 10, 18, 24, 42–45], parotid glands[18, 4, 10, 33, 24, 44, 7, 40, 11, 21, 46, 15, 16], submandibular glands[4, 24, 46], pituitary gland[9], thyroid[13, 41, 47], eye globes[9, 13, 14, 17, 31, 40, 41], eye lenses[13, 14, 16, 17, 41], optic nerves[5, 9, 14, 26, 31, 41], optic chiasm[5, 9, 16, 26, 31], larynx[40, 46], pharyngeal constrictor muscle[46, 48], pterygoid muscles[4], tongue muscles[28] and lymph nodes[4, 6, 12, 18–20, 24, 25, 29, 30, 32, 49]. Despite the considerable attention, the demonstrated results are still not satisfactory for the clinical usage and automated methods cannot accurately segment OARs in the presence of tumors and severe pathologies.

In the work, we propose the first deep learning-based algorithm for HaN CT image segmentation. Deep learning techniques such as convolutional neural networks (CNNs) have demonstrated impressive performance in computer vision and medical image analysis applications[50]. Being supervised, CNNs first study the appearance of objects of interest in a training set of segmented images. When a previously unseen test image arrives, CNNs mark the voxels that have appearance similar to the corresponding OARs. The performance of CNNs was validated on segmentation of spinal cord, mandible, parotid glands, submandibular glands, larynx, pharynx, eye globes, optic nerves and optic chiasm. The paper is organized as follows. Section 2 presents the theoretical concept of CNNs and the way we used CNNs for image segmentation. In Section 3, we described a validation database of HaN CT images and the results of CNN-based segmentation. We compared the results with state-of-the-art in Section 4, and concluded in Section 5.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

2. Methods

2.1. Convolutional neural networks

Convolutional neural network is a special kind of multi-level perceptron architecture, where an input image passes through a sequence of classification tests that can extract and recognize its consistent intensity patterns and finally make a prediction about the image according to these patterns. In contrast to alternative machine learning-based algorithms, CNNs take spatial information into account so that neighboring pixels are analyzed together. This behavior combined with abilities for generalization makes CNNs superior to other approaches on a number of computer vision applications, such as handwritten text recognition, face detection and object classification[51]. A CNN is organized as a set of different layers, where output of a current layer is used as an input for the next layer. A convolution layer – the main component of CNNs – serves to generate a predefined number of rectangular features computed from the output of the previous layer or the original image, if this convolution layer is positioned at the beginning of a CNN. As a results, convolution layers extract elementary features of the input such as edges and corners, which after passing through next layers are combined into more complex high-order features. A convolutional layer is usually followed by a rectified linear units (ReLU), hyperbolic tangent or sigmoid layer that introduces non-linearity to the convolution layer response and reduces overfitting and vanishing or exploding gradient effects. A pooling layer then partitions the input into non-overlapping rectangles and returns the maximum or average value of each rectangle. This makes CNNs more robust against local transformations of the input as pooling layers are invariant to small shift, rotation and scaling. Moreover, pooling considerably reduces the input size and consequently the number of network parameters, and therefore prevents CNN overfitting. Standard networks composed of convolution-ReLU-pooling layer sets are sometimes sensitive to initialization and require fine parameters tuning in order to avoid gradient explosions. The recent study has demonstrated that batch-based normalization (bNorm) of the each convolution layer makes networks less sensitive to input parameter initialization and more than 14-times increases network convergence[52]. Finally, the robustness of the network can be improved by random removal of some neurons, which is performed using dropout layers[53]. After a sequence of convolution-bNorm-ReLU-dropout-pooling layer sets, the input image shrinks into a small set of high-order features and finally passes through fully connected layers that generate the network output (Fig. 1).

In order to use CNNs for image annotation, let us have a training set T of 3D images with an object (objects) of interest annotated. An image patch of each voxel that belongs to the object of interest contains some intensity patterns that make it distinguishable from patches of other voxels in the image. The goal of a CNN is to correctly capture these patterns, and use them to identify object voxels in unseen images. As it was shown in the recent studies, CNNs trained on a set of tri-planar patches, i.e. square intensity neighborhoods extracted at three orthogonal cross-sections and centered at the target voxel, exhibit similar segmentation performance but dramatic acceleration in comparison to CNNs trained on 3D patches[54]. We therefore designed a tri-planar patch-based network with three sets of convolution-bNorm-ReLU-dropout-pooling layers, a fully connected layer and a softmax layer that converts network output into classification prediction values (Fig. 1).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

2.2. Detection of organs-at-risks

In the training set T of HaN CT images, experienced clinicians defined objects of interest as OARs including spinal cord, mandible, parotid glands, submandibular glands, eye globes, optic nerves, optic chiasm, larynx, and pharynx. The goal is to detect and segment these OARs in a previously unseen target image. Although initial OARs detection is not mandatory, it considerably reduces segmentation computational costs as only voxels that may geometrically belong to OARs are analyzed. For this aim, we automatically detect the patient's head, which is used as a reference point for approximation of OARs positions. Following the geometry of the human head, we observe that the image gradients computed at skull boundaries are, in general, oriented towards the brain center, which allow us to apply the existing algorithm developed for detection of spherical structures, such as femoral heads[55]. For every voxel $\mathbf{x}$ with gradient magnitude $|\nabla I_{\mathbf{x}}|$ above a certain threshold, indicating that voxel $\mathbf{x}$ may belong to skull surface, we compute both positive $\nabla I_{\mathbf{x}}$ and negative $-\nabla I_{\mathbf{x}}$ gradient vectors. A high number of these gradient vectors intersect around the center of the skull, as a high number of gradients $\nabla I_{\mathbf{x}}$ and $-\nabla I_{\mathbf{x}}$ are oriented towards the head center (Fig. 2). We therefore search for the reference voxel where the number of intersected gradients is the highest, and define it as the head center. The regions of interest for each OAR are approximated according to this center point.

2.3. CNN-based segmentation of head and neck CT images

To segment each object on a previously unseen test image, the corresponding CNN must be trained to recognize the intensity appearance of voxels that belong to the object and distinguish it from the intensity appearance of surrounding background voxels. During the network training phase for a selected object of interest $\mathcal{S}$, we extracted set $\mathcal{P}_I^+$ of k voxels randomly sampled with repetitions from the set of voxels that belong to the binary mask that represents object $\mathcal{S}$ in image $I \in T$. Set $\mathcal{P}_I^-$ of the same cardinality of k voxels is randomly sampled with repetitions from voxels that do not belong to object $\mathcal{S}$ but located not closer than r mm from the nearest object voxel in image $I \in T$. We do not consider voxels that are located further than R mm from object $\mathcal{S}$ because its position is restricted according to the patient anatomy and imaging protocol, and very distant voxels will therefore not be analyzed during segmentation and cannot enrich the corresponding CNN. For each voxel from $\mathcal{P}_I^+ \left(\mathcal{P}_I^-\right)$, we extract tree orthogonal patches from image I that define a positive (negative) training sample. A complete set of $k \cdot |T|$ positive and $k \cdot |T|$ negative training samples, generated for all images from T, is used by CNN for modeling the intensity appearance pattern that separates object $\mathcal{S}$ from background[56] (Fig. 3). The trained network is next applied to classify voxels in the region of interest of the test image where the corresponding OARs are expected to be located. The classification result of the network computed for a voxel $\mathbf{x}$ passes through the softmax layer to obtain the probability $p_{\mathbf{x}}^{\mathcal{S}}$ that voxel $\mathbf{x}$ belongs to object $\mathcal{S}$.

2.4. Markov random fields and segmentation post-processing

Segmentation with CNNs usually produces smooth probabilities $p_{\mathbf{x}}^{\mathcal{S}}$ because voxels located close to each other have similar intensity appearance. However, CNNs do not explicitly

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

incorporate voxel connectivity information and the morphology of OARs. As the morphology of OARs strongly depends on the location of the tumor and often poorly predictable, we rely on connectivity information and apply Markov random fields (MRF) to improve CNN results. According to the MRF formulation, the target 3D image can be described as a graph $\mathscr{G} = \{\mathscr{V}, \boldsymbol{\varepsilon}\}$, where vertices $\mathscr{V}$ correspond to image voxels, and edges $\boldsymbol{\varepsilon}$ connect vertices that correspond to neighboring voxels. For each object $\mathscr{S}$, each vertex can be assigned label 1 if it belongs to $\mathscr{S}$, and assigned 0 otherwise. The label of a voxel depends on its similarity to object $\mathscr{S}$, i.e. probability $p_{\mathbf{x}}^{\mathscr{S}}$, and similarity to object $\mathscr{S}$ of its neighbors:

$$L^{\mathscr{S}}=\arg\min_{l_{\mathbf{x}}}\left(\sum_{\mathbf{x}\in V}\phi\left(p_{\mathbf{x}}^{\mathscr{S}},l_{\mathbf{x}}\right)+\sum_{(\mathbf{x},\mathbf{y})\in\varepsilon}\psi_{\mathbf{x},\mathbf{y}}\Delta\left(l_{\mathbf{x}},l_{\mathbf{y}}\right)\right), \qquad (1)$$

where function $\phi\left(p_{\mathbf{x}}^{\mathscr{S}},l_{\mathbf{x}}\right)=(l_{\mathbf{x}}{==}1)\,?1-p_{\mathbf{x}}^{\mathscr{S}}:p_{\mathbf{x}}^{\mathscr{S}}$ defines the cost of assigning label 1 to voxel b as $1-p_{\mathbf{x}}^{\mathscr{S}}$, and the cost of assigning label 0 to voxel **x** as $p_{\mathbf{x}}^{\mathscr{S}}$. Function $\psi_{\mathbf{x},\mathbf{y}}=\exp\left(-\frac{\left(p_{\mathbf{x}}^{\mathscr{S}}-p_{\mathbf{y}}^{\mathscr{S}}\right)^2}{2\sigma^2}\right)$ measures similarity between probabilities $p_{\mathbf{x}}^{\mathscr{S}}$ and $p_{\mathbf{y}}^{S}$ of voxels **x** and **y**, respectively, and $\Delta(l_{\mathbf{x}}, l_{\mathbf{y}})$ is an indicator function that equals 0 if labels $l_{\mathbf{x}} == l_{\mathbf{y}}$, and equals 1 otherwise. The first term of Equation 1 tries to assign voxel labels according to the probabilities $p_{\mathbf{x}}^{\mathscr{S}}$ computed by the CNN, whereas the second term tries to make labels of all neighboring voxels, i.e. labels of all voxels in the image, the same. Searching for the equilibrium between these two terms usually makes the resulting label mask smooth and suppresses both small groups of voxels with high probabilities $p_{\mathbf{x}}^{\mathscr{S}}$ surrounded by voxels with low probabilities $p_{\mathbf{x}}^{\mathscr{S}}$, and vice versa. For solving this MRF, we used a publicly available software FastPD, which can find the optimal MRF-based labeling almost in real time[57].

Despite its effectiveness, MRF approach cannot suppress large groups of isolated voxels, and fill large cavities in resulting label maps L^{S}. However, we know in advance that OARs represent solid connected structures with smooth boundaries, and should not be separated into several components and do not have internal cavities. We therefore extract the largest connected component of voxels L^{S} so that noise voxel groups with probability $l_{\mathbf{x}} > 0.5$ are removed. A sequence of dilation operations is then used to fill all the cavities in the obtained largest component. The same number of erosion operations compensate for the overexpantion of the dilated component, which finally defines segmentation of object $\mathscr{S}$ in the test image. After performing CNN-based voxel classification, MRF-based smoothing, small component removal and dilate-erode operations for all objects of interest, segmentation of the test image is complete.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3. Experiments and Results

3.1. Image database

We collected 3D CT images of 50 patients scheduled for head and neck radiotherapy. All images were axially reconstructed and had in-plane resolution between 0.781 and 1.310 mm with the scan matrix of 512×512, whereas the slice thickness varied between 1.5 and 2.5 mm. Experienced radiologists annotated mandible in 48 images; optic nerves and right eye globe in 47 images; left parotid, left eye globe and chiasm in 46 images; spinal cord, right parotid and larynx in 45 images; pharynx in 39 images; and both submandibular glands in 31 images, which rounds to 563 objects to segment. Inclusion of certain OARs into the treatment planning procedure depended on the position of the tumor and image field of view.

3.2. Validation

We performed five-folder cross-validation, where, for example, for segmenting the image with number 15, images with numbers 1-10 and 21-50 were used for CNN training. During the training phase, all voxels than belong to the reference segmentations of OARs were extracted. For each OAR mask and its surrounding structures, we randomly subsampled $k = 25000$ positive and $k = 25000$ negative patches of 27×27×3 size, i.e. axial, sagittal and coronal patches of 27×27 mm, that formed the training set of samples for the corresponding image. The negative samples were located not closer than $r = 2$ mm and not further than $R = 40$ mm from the surface of the corresponding OARs. According to the positive and negative samples from all training images, the networks were optimized using the stochastic gradient descent scheme, initialized with a learning rate of 0.0005 and momentum of 0.9. Addition gradient clipping procedure is used to prevent gradient explosions, which sometimes occur during CNN training. The training samples were separated into batches with 250 samples each, and all batches were analyzed in 25 epochs, i.e. training procedure runs 25 times through all training samples. The binary masks obtained after performing CNN-based voxel classification, small component removal and dilate-erode operations were compared to reference segmentation masks by computing true positive (TP) rate that is the number voxels classified as the target OAR in both reference and automated segmentation, false positive (FP) rate that is the number of voxels mis-classified as the target OAR by automated segmentation, and false negative (FN) rate that is the number of OAR voxels mis-classified as background by automated segmentation. In the case of spinal cord segmentation, we did not consider automatic segmentation on the slices depicting thoracic spine where the spinal cord is not segmented by the clinicians. The intuition is that the reference segmentation of spinal cord does not continue through the whole vertebral column, and stops at the slice where the cord is not anymore affected during the treatment. At the same time, automatic segmentation continues segmenting spinal cord beyond this slice, which does not harm the treatment planning and shall not be straightforwardly considered as an error. The segmentation error was numerically measured using Dice coefficient (DSC) = $2TP/(2TP + FN + FP)$. The mean DSC varied from 37.4% for chiasm segmentation to 89.5% for mandible segmentation (Table 1 and Figs. 4 and 5).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3.3. Computation performance

The proposed algorithm uses two publicly available toolboxes: Computational Network Toolkit (CNTK)[56] and FastPD[57], and a set of functions for pre- and post-processing, which we implemented in C++ and executed on a personal computer with Intel Core i7 processor at 4.0 GHz and 32 GB of memory. A graphical processing unit (GPU) was used by CNTK for deep learning-based segmentation. The training of the networks took 30 minutes per OARs, which rounds to 6.5 hours for training 13 networks for all OARs. The analysis of one image took around 4 minutes, where only 1.5 minutes were used for deep learning classification of testing patches, 2.5 minutes were used for reading-writing of images and samples, OARs detection, image transformation and post-processing operations, whereas MRF refinement required less than a second.

4. Discussion

We presented the first attempt to use CNNs for segmentation of OARs from HaN CT images. As a number of automated and semi-automated methods have been proposed to address this problem, we compared our results with the current state-of the-art and devised conclusions about the potential of using CNN for HaN radiotherapy planning. Segmentation of an OAR strongly depends on its size, shape, clarity of boundaries, presence of pathologies and overall visibility in the CT image modality. The comprehensive comparison should be based not only on the results of existing automated segmentation methods but also on the performance of commercially available software and inter-observer variability. We summarize research published on the topic of HaN OARs segmentations, compare the reported results with the proposed framework performance and identify baselines and milestones for individual OARs. We also present the inter-observer variability for some OARs reported in the literature, which is important for estimation of the complexity of the segmentation problem. However, it is important to mention that this is not the direct comparison obtained on the same database, but the results of alternative automated methods, commercial software and human performance reported on different databases.

4.1. Spinal cord segmentation

Radiation therapy can cause toxicity to the nervous system, if spinal cord is affected during the treatment. This can lead to self-limited transient or progressive myelopathy, or, in rare cases, to lower motor neuron syndrome. As spinal cord is poorly visible in CT images, clinicians usually segment the region encompassed by vertebral foramens that ensures some safe margins for spinal cord. Although surface of cervical vertebrae is a guide for segmentation, identification of the exact boundaries of spinal cord in the test CT image is a challenging task. Existing automated segmentation methods reported Dice coefficient of 62%[16], 75%[16], 76%[41], 78%[4], 83%[18] and 85%[13]. La Macchia et al.[45] measured the performance of commercial software on spinal cord segmentation and obtained Dice coefficient of 70%, 81% and 78% for VelocityAI 2.6.2, MIM 5.1.1 and ABAS 2.0 systems, respectively. To objectively estimate the complexity of the problem, it is important to mention that inter-observer variability was reported to be of 77%[48], 79%[40, 41], 80%[58] DSC, 60%[59] in terms of conformity level, and 78%[60], and 89%[61] in terms of maximal volume difference. We can therefore conclude that the proposed CNN-based spinal cord

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

segmentation result of 87.0 ± 3.2% DSC compares favorably against the existing segmentation algorithms, commercial software and inter-observer variability (Fig. 5). This achievement can be explained by the fact that spinal cord has a very consistent intensity pattern that can be accurately modeled by machine learning approaches such as CNNs. On the other hand, atlas-based algorithms are often too restricted by the registration regularization, so that the resulting transformation field is smooth and not distorted but, at the same time, atlasing cannot perfectly capture variable curvature of the spine, cervical vertebra sizes, etc.

4.2. Mandible segmentation

Irradiation of dental area can cause dental decay, postextraction osteoradionecrosis and implant failure, therefore mandible has to be spared during the HaN radiotherapy procedure[43]. Mandible identification is facilitated by its relatively large size and rigid shape, whereas the fact that mandible has a better contrast in CT images in comparison to the surrounding soft structures simplifies its segmentation. The main challenge is the presence of severe image artifacts around teeth due to metallic dental restorations. These artifacts hamper correct identification of mandible borders and often corrupt the appearance of the surrounding structures such as tongue, parotid and submandibular glands. These artifacts were, in general, excluded from mandible segmentation, as CNNs were able to correctly model the composition of dark and bright voxel groups that usually characterize dental artifacts in CT images. Moreover, the teeth region where dental artifacts are manifested was not included into manual segmentation, which additionally supported automated segmentation. Automated mandible segmentation has received considerable attention in the research community, and the obtained results measured in terms of Dice coefficient were of 78%[62], 80%[4], 82%[10, 44], 86%[18], 90%[10] and 93%[24, 42] whereas existing commercial VelocityAI 2.6.2, MIM 5.1.1 and ABAS 2.0 software systems demonstrated Dice coefficient of 84%, 86% and 89%, respectively. The proposed CNN-based algorithm segmented mandible with 89.5 ± 3.6% DSC, which is comparable with the best performing segmentation algorithms (Fig. 5, 2nd and 3rd rows). Moreover, the demonstrated results are similar to the semi-automatic segmentations 89%[45] DSC, and inter-observer variability of 87%[58] and 89%[40] DSC.

4.3. Parotid and submandibular gland segmentation

Xerostomia – changes in volume, consistency and pH of saliva – is one of the main complication after HaN radiation therapy[63]. As around 60% and 30% of saliva is produced by parotid or submandibular glands, xerostomia is usually manifested when these glands are irradiated during treatment[63]. Segmentation of parotid or submandibular glands in CT images is challenging due to their irregular shapes, poorly visible boundaries that are additionally corrupted by dental artifacts. According to literature overview, automated segmentation results of 57%[16], 65%[16], 68%[10] 72%[7], 77%[33], 79%[40, 44], 80%[4, 18, 46], 83%[4, 15, 24], 84%[11], 85%[10] DSC for parotid glands, and 70%[4], 80%[46] and 83%[24] DSC for submandibular glands were reported. At the same time, VelocityAI 2.6.2, MIM 5.1.1 and ABAS 2.0 software systems demonstrated parotid gland segmentation performance of 73%, 79% and 79% DSC, respectively[45]. The agreement among radiologists is slightly better than between automated and manual segmentation and is reported to be of 76%[58], 79%[40], 85%[33]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

and 90%[46] DSC, and 77%[60, 61] volume difference for parotid glands, and 91%[46] DSC and 77%[60] volume difference of submandibular glands. The results of the proposed algorithm are of 77.3 ± 5.8% DSC for both parotid glands, which is comparable with the average results of the existing segmentation methods and above average in comparison to the performance of the commercial software (Fig. 5). Manual segmentation results indicate that parotid glands may be challenging to annotate even for a human. Moreover, parotid glands are often affected by pathologies, which considerably reduces the reliability of segmentation. The results of the proposed algorithm are of 71.4 ± 11.6% DSC for submandibular glands, which is lower than the average performance of existing algorithms, and commercial software (Fig. 4, 2nd and 3rd rows). There are two reasons why CNNs are not as effective as atlas-based segmentation for submandibular glands. First, the glands do not have very distinguishable intensity features that would make them different from the surrounding structures in contrast to spinal cord and mandible. Second, atlasing deformation is relatively restricted, and therefore, even if a gland cannot be perfectly recognized, atlasing can still roughly identify the gland positioning. At the same time, CNN relies purely on image intensities around glands, and therefore may sometimes get confused. The positive consequence of this property is that segmentation of abnormal submandibular gland is going to be more challenging for atlasing, as it cannot efficiently incorporate the tumor position. We also observe that submandibular gland segmentation accuracy strongly depends on the particular image database, and even human performance can vary dramatically.

4.4. Larynx and pharynx

Dysphagia or swallowing dysfunction is considered to be the second most commonly occurring complication of the HaN radiation therapy[63]. The manifestation of dysphagia has been shown to be consistently caused by irradiation of supraglottic larynx and pharyngeal constrictor muscles[63]. Although segmentation of larynx and pharynx is often required for treatment planning, automated segmentation of them has not been receiving as much attention as segmentation of spinal cord, mandible and glands, due to highly variable intensity of larynx and poorly defined boundaries and complex shape of pharynx in CT images. To the best of our knowledge, the existing studies on the topic reported automated segmentation results of 58%[46] and 73%[48] DSC for larynx and 50%[46] and 64%[48] DSC for pharyngeal constrictor muscles. Commercial VelocityAI 2.6.2, MIM 5.1.1 and ABAS 2.0 software systems demonstrated better performance of 81%, 83% and 84% DSC, respectively, on larynx segmentation, and 63%, 66% and 65% DSC, respectively, on pharynx segmentation[45]. Clinicians demonstrated an agreement of 60%[48] and 83%[40] DSC for larynx segmentation. We obtained segmentation results of 85.6 ± 4.2% DSC for larynx and 69.3 ± 6.3% DSC for pharynx that is superior or comparable to the best performing automated segmentation and inter-observer variability (Fig. 5, 2nd-4th rows). Designed CNNs were therefore able to correctly capture the appearance of laryngeal cartilages and pharyngeal area that is surrounded by cervical spine and airways.

4.5. Eyes, optical nerves and chiasm

In the case of nasopharyngeal cancer, tumors can be located in the close proximity to human vision system, irradiation of which can cause serious complications such as radiation-induced optic neuropathy[64]. Modern HaN radiotherapy protocols therefore require pre-

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

operative annotation of eye globes, optical nerves and chiasm. Eyes have well-defined spherical shape and, despite poor visibility of the globe surface in CT images, can be accurately segmented with 80%[13, 41] and 81%[9] DSC using atlas- and model-based algorithms. The optic nerves, which are thin elongated structures that start at the eyes and meet below the hypothalamus, can be automatically segmented with 38%[9], 58%[41], 42%[13] and 61%[13] DSC. The optic nerve intersection, i.e. optic chiasm, has a very small size and poorly visible boundaries and can be segmented automatically with relatively low accuracy of 39%[26] and 41%[9] DSC. Similarly to the results of automated segmentation, inter-observer agreement is higher for eye globes, namely 86%[26, 41] and 89%[40] DSC, lower for optic nerves, namely 51%[26], 57%[48] and 60%[41] DSC, and the lowest for optic chiasm, namely 38%[48] and 41%[26] DSC. The proposed CNN-based algorithm with 88.0 ± 3.2% DSC outperforms the existing state-of-the-art algorithms for eye globe segmentation, with 64.2 ± 7.2% DSC similar to the best performing algorithm[13] for optic nerve segmentation, but with 37.4 ± 13.4% DSC, which is worse than the existing algorithms for chiasm segmentation (Fig. 5, 1st row).

4.6. Methodological comparison

As a supervised technique, CNNs considerably relies on the quality and representativeness of the training data. The richer is the training dataset, the more accurate and reliable is segmentation, which however does not affect computational time of CNN-based segmentation. In contrast, the computational complexity of multi-image atlasing linearly increases with \ the number of training images so that the target image may require hours to be segmented. The training phase of CNNs is relatively lengthy but happens only once and therefore does not affect segmentations of individual target images. For both CNNs and atlasing, segmentation accuracy is often lower for OARs located close to tumors. Indeed, the tumor presence changes the appearance and morphology of the OARs and the consistent appearance patterns learned by CNNs become less representative. This issue is more pronounced in the case of atlasing, as registering an image with no tumor at a specific location to an image with tumor at this location is extremely challenging. On the other hand, image artifacts and local intensity inhomogeneity are more destructive for CNNs than for atlasing. Similarly as in the case of tumor positioning, patterns learned by CNNs become less representative, whereas artifacts and intensity inhomogeneity does not affect the morphology of OARs, and therefore altasing is still able to perform registration.

Finally, we want to estimate the contribution of the MRF-based smoothing on framework performance. Having CNN-based classifications, MRF smoothen surface of the resulting binary masks and remove isolated mis-classified voxels or groups of voxels. The performance improvement was therefore lower for objects with clearly-visible boundaries, e.g. mandible segmentation was improved by 0.24 Dice, and well-defined shape, e.g. left and right eye globe segmentations were improved by 0.45 and 0.03 Dice, respectively. At the same time, CNNs produced irregular segmentation for poorly-visible left and right submandibular glands, therefore, the contribution of MRF was considerable and resulted in segmentation improvement by 3.73 and 3.93 Dice, respectively. In average, MRF increased the Dice coefficient by 0.91 for all OARs and visually improved segmentation boundaries.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5. Conclusion

In this paper, we presented the first attempt of using deep learning concept of convolutional neural networks for segmentation of OARs in the head and neck CT images. We observed that CNNs demonstrated performance superior or comparable to the state-of-the-art on segmentation of spinal cord, mandible, larynx, pharynx, eye globes and optic nerves, and inferior performance on segmentation of parotid glands, submandibular glands and optic chiasm. These results confirm that CNNs well-generalize the intensity appearance of objects with recognizable boundaries, e.g. spinal cord, mandible, pharynx and eye globes. For the objects with poorly recognizable boundaries, e.g. submandibular glands and optic chiasm, additional information may be required for CNN-based segmentation. Following the clinical procedure, magnetic resonance images, which have better image contrast of soft tissues than CT image, can be included as an information source. However, this inclusion requires accurate multi-modal registration of magnetic resonance to CT images, which is a very non-trivial task when images have different fields of view and resolution. Combining CT and magnetic resonance images may have potential but is not straightforward. In general, we confirm a high potential of deep learning for HaN image segmentation.

Acknowledgments

This work was supported by a Faculty Research Award from Google Inc. and research grants from the National Institutes of Health (1R01CA176553 and R01E0116777).

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65(2):87–108. [PubMed: 25651787]
2. 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. 2010; 77(3):950–958. [PubMed: 20378266]
3. Sharp G, et al. Vision 20/20: perspectives on automated image segmentation for radiotherapy. Med Phys. 2014; 41(5):50902.
4. Han X, et al. Atlas-based auto-segmentation of head and neck CT images. Med Image Comput Comput-Assist Interv MICCAI Int Conf Med Image Comput Comput-Assist Interv. 2008; 11(Pt 2): 434–441.
5. Commowick O, Grégoire V, Malandain G. Atlas-based delineation of lymph node levels in head and neck computed tomography images. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2008; 87(2): 281–289.
6. Voet PWJ, Dirkx MLP, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJM. Does atlas-based autosegmentation of neck levels require subsequent manual contour editing to avoid risk of severe target underdosage? A dosimetric analysis. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2011; 98(3):373–377.
7. Daisne JF, Blumhofer A. Atlas-based automatic segmentation of head and neck organs at risk and nodal target volumes: a clinical validation. Radiat Oncol Lond Engl. 2013; 8:154.
8. Bondiau PY, et al. Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. Int J Radiat Oncol Biol Phys. 2005; 61(1):289–298. [PubMed: 15629622]
9. Isambert A, et al. Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2008; 87(1):93–99.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

10. Sims R, et al. A pre-clinical assessment of an atlas-based automatic segmentation tool for the head and neck. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2009; 93(3):474–478.
11. Fritscher KD, Peroni M, Zaffino P, Spadea MF, Schubert R, Sharp G. Automatic segmentation of head and neck CT images for radiotherapy treatment planning using multiple atlases, statistical appearance models, and geodesic active contours. Med Phys. 2014; 41(5):51910.
12. Teguh DN, et al. Clinical validation of atlas-based auto-segmentation of multiple target volumes and normal tissue (swallowing/mastication) structures in the head and neck. Int J Radiat Oncol Biol Phys. 2011; 81(4):950–957. [PubMed: 20932664]
13. Fortunati V, et al. Tissue segmentation of head and neck CT images for treatment planning: a multiatlas approach combined with intensity modeling. Med Phys. 2013; 40(7):71905.
14. Verhaart RF, et al. The relevance of MRI for patient modeling in head and neck hyperthermia treatment planning: a comparison of CT and CT-MRI based tissue segmentation on simulated temperature. Med Phys. 2014; 41(12):123302. [PubMed: 25471984]
15. Wachinger C, Fritscher K, Sharp G, Golland P. Contour-Driven Atlas-Based Segmentation. IEEE Trans Med Imaging. 2015; 34(12):2492–2505. [PubMed: 26068202]
16. Hoang Duc AK, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. Med Phys. 2015; 42(9):5027–5034. [PubMed: 26328953]
17. Fortunati V, Verhaart RF, Niessen WJ, Veenland JF, Paulides MM, van Walsum T. Automatic tissue segmentation of head and neck MR images for hyperthermia treatment planning. Phys Med Biol. 2015; 60(16):6547–6562. [PubMed: 26267068]
18. Zhang T, Chi Y, Meldolesi E, Yan D. Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy. Int J Radiat Oncol Biol Phys. 2007; 68(2):522–530. [PubMed: 17418960]
19. Chen A, Deeley MA, Niermann KJ, Moretti L, Dawant BM. Combining registration and active shape models for the automatic segmentation of the lymph node regions in head and neck CT images. Med Phys. 2010; 37(12):6338–6346. [PubMed: 21302791]
20. Teng CC, Shapiro LG, Kalet IJ. Head and neck lymph node region delineation with image registration. Biomed Eng OnLine. 2010; 9:30. [PubMed: 20569461]
21. Yang X, et al. Automated segmentation of the parotid gland based on atlas registration and machine learning: a longitudinal MRI study in head-and-neck radiation therapy. Int J Radiat Oncol Biol Phys. 2014; 90(5):1225–1233. [PubMed: 25442347]
22. Schreibmann E, Xing L. Image registration with auto-mapped control volumes. Med Phys. 2006; 33(4):1165–1179. [PubMed: 16696494]
23. Chao M, Xie Y, Moros EG, Le QT, Xing L. Image-based modeling of tumor shrinkage in head and neck radiation therapy. Med Phys. 2010; 37(5):2351–2358. [PubMed: 20527569]
24. Qazi AA, Pekar V, Kim J, Xie J, Breen SL, Jaffray DA. Auto-segmentation of normal and target structures in head and neck CT images: a feature-driven model-based approach. Med Phys. 2011; 38(11):6160–6170. [PubMed: 22047381]
25. Commowick O, et al. An efficient locally affine framework for the smooth registration of anatomical structures. Med Image Anal. 2008; 12(4):427–441. [PubMed: 18325825]
26. Deeley MA, et al. Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. Phys Med Biol. 2011; 56(14): 4557–4577. [PubMed: 21725140]
27. Leavens C, et al. Validation of automatic landmark identification for atlas-based segmentation for radiation treatment planning of the head-and-neck region. SPIE Med Imaging. 2008
28. Ibragimov B, et al. Segmentation of tongue muscles from super-resolution magnetic resonance images. Med Image Anal. 2015; 20(1):198–207. [PubMed: 25487963]
29. Yan J, Zhuang T, Zhao B, Schwartz LH. Lymph node segmentation from CT images using fast marching method. Comput Med Imaging Graph Off J Comput Med Imaging Soc. 2004; 28(1–2): 33–38.
30. Dornheim J, Seim H, Preim B, Hertel I, Strauss G. Segmentation of neck lymph nodes in CT datasets with stable 3D mass-spring models segmentation of neck lymph nodes. Acad Radiol. 2007; 14(11):1389–1399. [PubMed: 17964462]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

31. Bekes G, Máté E, Nyúl LG, Kuba A, Fidrich M. Geometrical model-based segmentation of the organs of sight on CT images. Med Phys. 2008; 35(2):735–743. [PubMed: 18383695]

32. Gorthi S, et al. Segmentation of Head and Neck Lymph Node Regions for Radiotherapy Planning Using Active Contour-Based Atlas Registration. IEEE J Sel Top Signal Process. 2009; 3(1)

33. Faggiano E, et al. An automatic contour propagation method to follow parotid gland deformation during head-and-neck cancer tomotherapy. Phys Med Biol. 2011; 56(3):775–791. [PubMed: 21239848]

34. Ibragimov B, Likar B, Pernus F, Vrtovec T. A game-theoretic framework for landmark-based image segmentation. IEEE Trans Med Imaging. 2012; 31(9):1761–1776. [PubMed: 22692901]

35. Ibragimov B, Likar B, Pernuš F, Vrtovec T. Shape representation for efficient landmark-based segmentation in 3-d. IEEE Trans Med Imaging. 2014; 33(4):861–874. [PubMed: 24710155]

36. Korez R, Ibragimov B, Likar B, Pernuš F, Vrtovec T. A Framework for Automated Spine and Vertebrae Interpolation-Based Detection and Model-Based Segmentation. IEEE Trans Med Imaging. 2015; 34(8):1649–1662. [PubMed: 25585415]

37. Yao J, et al. A multi-center milestone study of clinical vertebral CT segmentation. Comput Med Imaging Graph Off J Comput Med Imaging Soc. 2016; 49:16–28.

38. Wang CW, et al. Evaluation and Comparison of Anatomical Landmark Detection Methods for Cephalometric X-Ray Images: A Grand Challenge. IEEE Trans Med Imaging. 2015; 34(9):1890–1900. [PubMed: 25794388]

39. Wang CW, et al. A benchmark for comparison of dental radiography analysis algorithms. Med Image Anal. 2016; 31:63–76. [PubMed: 26974042]

40. Mattiucci GC, et al. Automatic delineation for replanning in nasopharynx radiotherapy: what is the agreement among experts to be considered as benchmark? Acta Oncol Stockh Swed. 2013; 52(7): 1417–1422.

41. Verhaart RF, Fortunati V, Verduijn GM, van Walsum T, Veenland JF, Paulides MM. CT-based patient modeling for head and neck hyperthermia treatment planning: manual versus automatic normal-tissue-segmentation. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2014; 111(1):158–163.

42. Pekar V, Allaire S, Kim J, Jaffray DA. Head and Neck Auto-Segmentation Challenge. MIDAS J. 2009

43. Thariat J, et al. Dentalmaps: Automatic Dental Delineation for Radiotherapy Planning in Head-and-Neck Cancer. Int J Radiat Oncol. 2012; 82(5):1858–1865.

44. Peroni M, et al. Automatic segmentation and online virtual CT in head-and-neck adaptive radiation therapy. Int J Radiat Oncol Biol Phys. 2012; 84(3):e427–433. [PubMed: 22672753]

45. La Macchia M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. Radiat Oncol Lond Engl. 2012; 7:160.

46. Thomson D, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. Radiat Oncol. 2014; 9:173. [PubMed: 25086641]

47. Chen A, Niermann KJ, Deeley MA, Dawant BM. Evaluation of multiple-atlas-based strategies for segmentation of the thyroid gland in head and neck CT images for IMRT. Phys Med Biol. 2012; 57(1):93–111. [PubMed: 22126838]

48. Tao CJ, et al. Multi-subject atlas-based auto-segmentation reduces interobserver variation and improves dosimetric parameter consistency for organs at risk in nasopharyngeal carcinoma: A multi-institution clinical study. Radiother Oncol. 2015; 115(3):407–411. [PubMed: 26025546]

49. Stapleford LJ, et al. Evaluation of automatic atlas-based lymph node segmentation for head-and-neck cancer. Int J Radiat Oncol Biol Phys. 2010; 77(3):959–966. [PubMed: 20231069]

50. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553):436–444. [PubMed: 26017442]

51. Goodfellow I, Bengio Y, Courville A. Deep Learning. 2016 Book in preparation for MIT Press.

52. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ArXiv150203167 Cs. 2015

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

53. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. J Mach Learn Res. 2014; 15:1929–1958.

54. Lai M. Deep Learning for Medical Image Segmentation. ArXiv150502000 Cs. 2015

55. Ibragimov B, Likar B, Pernuš F, Vrtovec T. Automated measurement of anterior and posterior acetabular sector angles. in. 2012:83151U–83151U–7.

56. Agarwal A, et al. An Introduction to Computational Networks and the Computational Network Toolkit. Microsoft Tech Rep. 2014 MSR-TR-2014-112.

57. Komodakis N, Tziritas G. Approximate Labeling via Graph Cuts Based on Linear Programming. IEEE Trans Pattern Anal Mach Intell. 2007; 29(8):1436–1453. [PubMed: 17568146]

58. Nelms BE, Tomé WA, Robinson G, Wheeler J. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer. Int J Radiat Oncol Biol Phys. 2012; 82(1):368–378. [PubMed: 21123004]

59. Mukesh M, et al. Interobserver variation in clinical target volume and organs at risk segmentation in post-parotidectomy radiotherapy: can segmentation protocols help? Br J Radiol. 2012; 85(1016):e530–536. [PubMed: 22815423]

60. Brouwer CL, et al. 3D Variation in delineation of head and neck organs at risk. Radiat Oncol Lond Engl. 2012; 7:32.

61. Geets X, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. Radiother Oncol J Eur Soc Ther Radiol Oncol. 2005; 77(1):25–31.

62. Tsuji SY, Hwang A, Weinberg V, Yom SS, Quivey JM, Xia P. Dosimetric Evaluation of Automatic Segmentation for Adaptive IMRT for Head-and-Neck Cancer. Int J Radiat Oncol. 2010; 77(3): 707–714.

63. Dirix P, Nuyts S. Evidence-based organ-sparing radiotherapy in head and neck cancer. Lancet Oncol. 2010; 11(1):85–91. [PubMed: 20129131]

64. Mayo C, Martel MK, Marks LB, Flickinger J, Nam J, Kirkpatrick J. Radiation dose-volume effects of optic nerves and chiasm. Int J Radiat Oncol Biol Phys. 2010; 76(3 Suppl):S28–35. [PubMed: 20171514]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

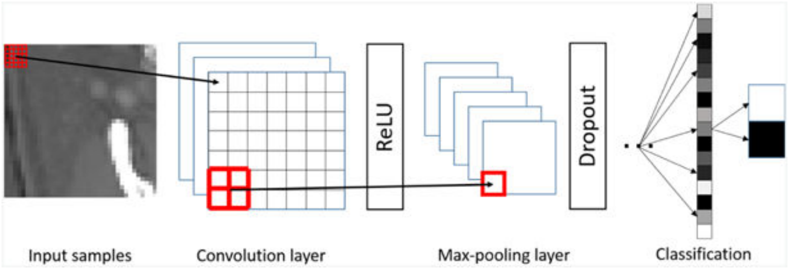


Figure 1.

A schematic illustration of the convolutional neural network architecture. Three orthogonal cross-sections around target voxel defines the input of the network that consists of three stacks of convolution, ReLU, max-pooling layer and dropout layers, fully-connected and softmax layers.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

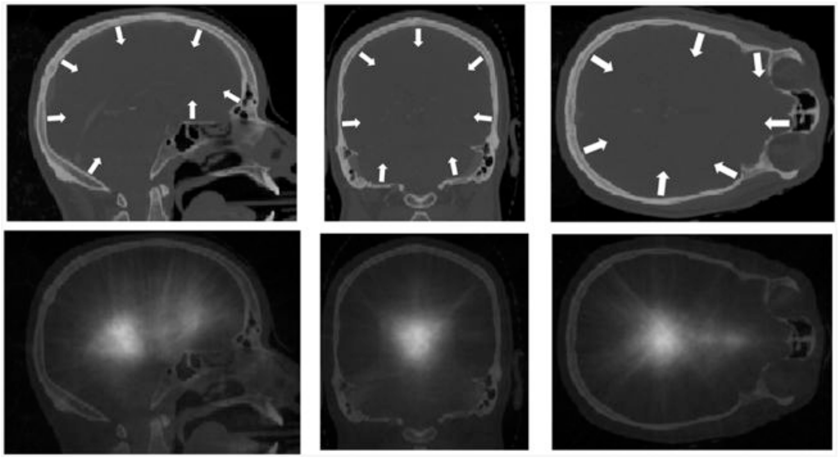


Figure 2.

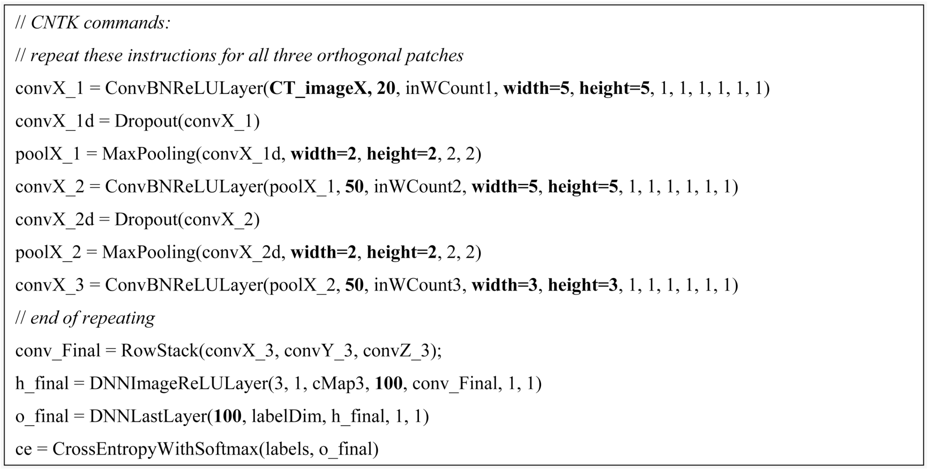
The strongest gradients of the target CT image (first raw) intersect around the center of the patient's skull (second raw).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



```
// CNTK commands:
// repeat these instructions for all three orthogonal patches
convX_1 = ConvBNReLULayer(CT_imageX, 20, inWCount1, width=5, height=5, 1, 1, 1, 1, 1, 1)
convX_1d = Dropout(convX_1)
poolX_1 = MaxPooling(convX_1d, width=2, height=2, 2, 2)
convX_2 = ConvBNReLULayer(poolX_1, 50, inWCount2, width=5, height=5, 1, 1, 1, 1, 1, 1)
convX_2d = Dropout(convX_2)
poolX_2 = MaxPooling(convX_2d, width=2, height=2, 2, 2)
convX_3 = ConvBNReLULayer(poolX_2, 50, inWCount3, width=3, height=3, 1, 1, 1, 1, 1, 1)
// end of repeating
conv_Final = RowStack(convX_3, convY_3, convZ_3);
h_final = DNNImageReLULayer(3, 1, cMap3, 100, conv_Final, 1, 1)
o_final = DNNLastLayer(100, labelDim, h_final, 1, 1)
ce = CrossEntropyWithSoftmax(labels, o_final)
```

Figure 3.

A schematic set of parameters and commands used to define convolutional neural network for segmentation of organ-at-risks in head-and-neck CT images. The parameters in bold correspond to the size of convolution and pooling layers.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

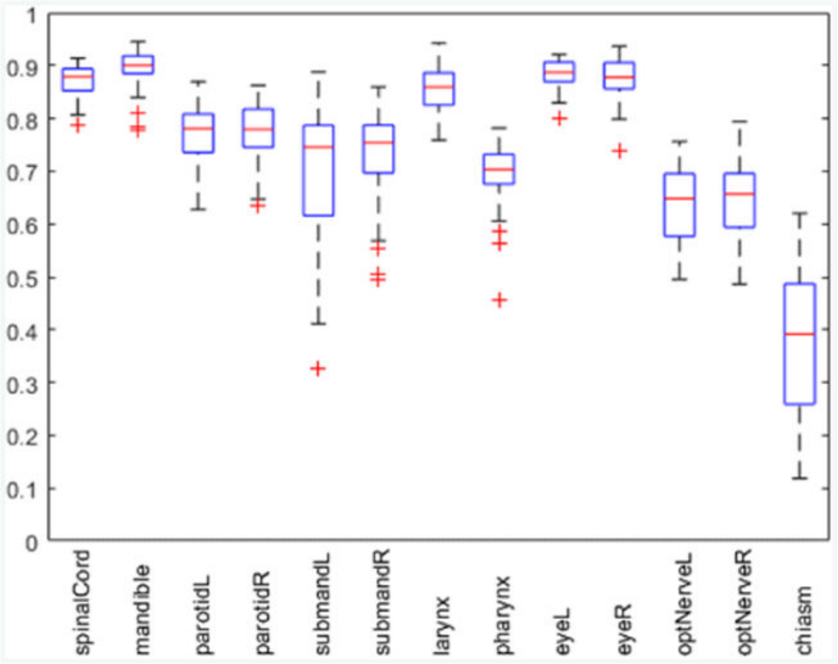


Figure 4.

Box plot results of convolutional neural network-based segmentation of organs-at-risks in head and neck CT images reported in terms of Dice coefficient.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

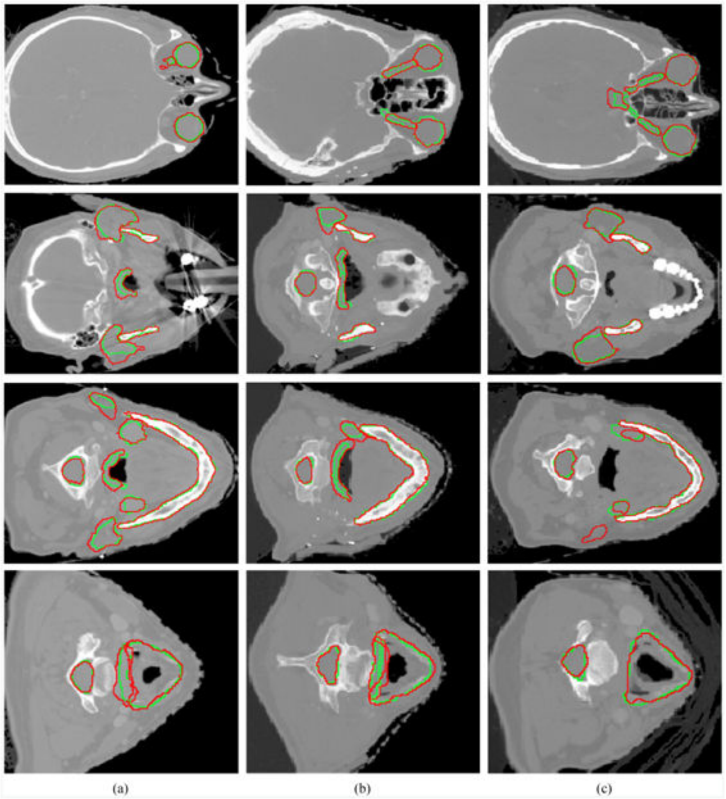


Figure 5.

Segmentation results for three (a-c) selected head and neck CT image, shown in four axial cross-sections. The reference segmentations are depicted in green and convolution neural network-based segmentations are depicted in red.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Results of convolutional neural network-based segmentation of organs-at-risks in head and neck CT images, given in terms of mean (± standard deviation) Dice coefficient (DSC).

Organ	**DSC (%)**
Spinal cord	87.0 ± 3.2
Mandible	89.5 ± 3.6
Parotid left	76.6 ± 6.1
Parotid right	77.9 ± 5.4
Submandibular left	69.7 ± 13.3
Submandibular right	73.0 ± 9.2
Larynx	85.6 ± 4.2
Pharynx	69.3 ± 6.3
Eye globe left	88.4 ± 2.7
Eye globe right	87.7 ± 3.7
Optic nerve left	63.9 ± 6.9
Optic nerve right	64.5 ± 7.5
Chiasm	37.4 ± 13.4