Intelligence-Based Medicine 5 (2021) 100026

Contents lists available at ScienceDirect

Intelligence-Based Medicine

journal homepage: www.journals.elsevier.com/intelligence-based-medicine

Auto-contouring FDG-PET/MR images for cervical cancer radiation therapy: An intelligent sequential approach using focally trained, shallow U-Nets

Atallah Baydoun [a], Ke Xu [b,c], Latoya A. Bethell [d], Feifei Zhou [d], Jin Uk Heo [d,e], Kaifa Zhao [b,f], Elisha T. Fredman [a], Rodney J. Ellis [g], Pengjiang Qian [b,c], Raymond F. Muzic Jr. [d,e,*], Bryan J. Traughber [g]

[a] Department of Radiation Oncology, University Hospitals Cleveland Medical Center, Cleveland, OH, USA
[b] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu, 214122, China
[c] Jiangsu Key Laboratory of Media Design and Software Technology, Jiangnan University, Wuxi, Jiangsu, 214122, China
[d] Department of Radiology, School of Medicine, Case Western Reserve University, Cleveland, OH, 44106, USA
[e] Department of Biomedical Engineering, Case Western Reserve University, Cleveland, OH, 44106, USA
[f] Department of Computing, Hong Kong Polytechnic University, Hung Hom, 999077, Hong Kong
[g] Department of Radiation Oncology, Penn State Cancer Institute, Hershey, PA, 17033, USA

ARTICLE INFO

Keywords:
Cervical cancer
Deep learning
Image segmentation
PET/MR Based radiation therapy
U-net

ABSTRACT

Background: Manual contouring for radiation therapy planning remains the most laborious and time consuming part in the radiation therapy workflow. Particularly for cervical cancer, this task is complicated by the complex female pelvic anatomy and the concomitant dependence on ^{18}F-labeled Fluorodeoxyglucose (FDG) positron emission tomography (PET) and magnetic resonance (MR) images. Using deep learning, we propose a new auto-contouring method for FDG-PET/MR based cervical cancer radiation therapy by combining the high level anatomical topography and radiological properties, to the low-level pixel wise deep-learning based semantic segmentation.

Materials/methods: The proposed method: 1) takes advantage of PET data and left/right anatomical symmetry, creating sub-volumes that are centered on the structures to be contoured. 2) Uses a 3D shallow U-Net (sU-Net) model with an encoder depth of 2.3) Applies the successive training of 3 consecutive sU-Nets in a feed forward strategy. 4) Employs, instead of the usual generalized dice loss function (GDL), a patch dice loss function (PDL) that takes into account the Dice similarity index (DSI) at the level of each training patch. Experimental analysis was conducted on a set of 13 PET/MR images was using a leave-one-out strategy.

Results: Despite the limited data availability, 5 anatomical structures - the gross tumor volume, bladder, anorectum, and bilateral femurs - were accurately (DSI = 0.78), rapidly (1.9 s/structure), and automatically delineated by our algorithm. Overall, PDL achieved a better performance than GDL and DSI was higher for organs at risk (OARs) with solid tissue (e.g. femurs) than for OARs with air-filled soft tissues (e.g. anorectum).

Conclusion: The presented workflow successfully addresses the challenge of auto-contouring in FDG-PET/MR based cervical cancer. It is expected to expedite the cervical cancer radiation therapy workflow in both, conventional and adaptive radiation therapy settings.

1. Introduction

Despite the current trends in aggressive screening and prevention, cervical cancer remains the fourth most common cancer among females after breast, colorectal, and lung cancers [1]. Cervical cancer is the fourth leading cause of cancer-related death among women worldwide [1], and the second-leading cause in the United States female population aged between 20 and 39 years [2]. The diagnostic approach is based on the combination of clinical, radiological, and tissue biopsy findings that determine the staging group as defined by the Fédération Internationale

* Corresponding author. 10900 Euclid Avenue, Biomedical Research Bldg, Floor 3, Cleveland, OH, 44106-4966, USA
E-mail addresses: atallah.baydoun@case.edu (A. Baydoun), 6171611028@stu.jiangnan.edu.cn (K. Xu), alexiabethell@case.edu (L.A. Bethell), feifei.zhou@case.edu (F. Zhou), jinuk.heo@case.edu (J.U. Heo), zhaokaifa@qq.com (K. Zhao), elisha.fredman@uhhospitals.org (E.T. Fredman), rellis1@pennstatehealth.psu.edu (R.J. Ellis), qianpjiang@jiangnan.edu.cn (P. Qian), raymond.muzic@case.edu (R.F. Muzic), bryan.traughber@case.edu (B.J. Traughber).

https://doi.org/10.1016/j.ibmed.2021.100026
Received 22 September 2020; Received in revised form 27 November 2020; Accepted 11 February 2021

de Gynécologie et d'Obstétrique (FIGO) [3]. The therapeutic strategy itself is grossly based on the FIGO staging, and often uses external beam radiation therapy (EBRT) with concurrent chemotherapy and brachytherapy boost [3,4]. In clinical practice, EBRT delivery requires first the acquisition of multi-modality imaging, followed by manual delineation of the contours to generate a radiation treatment dosimetric plan.

While staging and EBRT planning has been traditionally based on computed tomography (CT), magnetic resonance (MR) imaging has been advocated for a more thorough staging and treatment planning due to its high soft tissue contrast and ability to differentiate between the tumor and the normal cervix tissue [5]. Moreover, ^{18}F-labeled Fluorodeoxyglucose (FDG) positron emission tomography (PET) is recommended as it provides the most sensitive evaluation of the cancer spread to the pelvic and para-aortic lymph nodes [6]. As a result, MR and PET imaging were included in the guidelines as essential tools for the diagnosis and treatment planning of cervical cancer [6–8]. In regards to the MR acquisition parameters, gynecological societies recommend the acquisition of axial, Turbo Spin Echo (TSE), Single Shot (SSH) T2-weighted MR images for diagnostic and therapeutic guidance [8,9], as TSE-SSH technique reduces significantly the imaging time, eliminates geometric distortion, and decreases motion artifact compared to conventional spin echo sequences [10–12].

Prior to treatment planning, MR, PET, and CT images are co-registered. While CT images are mainly used to extract electron density information for dosimetric calculations, the radiation oncologist will depend essentially on MR and PET images to manually contour the targeted tumor volume and the organs at risk (OARs) defined as the important normal, non-cancerous structures close to or within the treatment field [13]. In the case of cervical cancer, OARs include the bladder, left femur, right femur, anorectum, and bowel [14]. While most OARs are usually well visualized with MR imaging alone, PET provides important complementary information for bladder, gross tumor volume (GTV) [13], and grossly involved lymph nodes requiring a radiation boost [15]. For example, GTV corresponds usually to the tumor area with an ^{18}F-FDG uptake higher than 2.5 mg/dL standard uptake value (SUV) or 20–40% of the maximum SUV [16], and the incorporation of PET in radiation therapy planning allowed the reduction of GTV inter-observer variability among radiation oncologists [17]. As for the bladder, it is also considered an area of high SUV, as ^{18}F-FDG is physiologically filtered through the kidney with little reabsorption [18].

Manual contouring remains the most burdensome, laborious task in radiation therapy planning [19,20] and is a major barrier for an accelerated automation of the cervical cancer radiation therapy procedure and for adaptive radiotherapy. In the case of cervical cancer: the planning is done on each individual axial slice and, assuming every structure is displayed over as many as 20 contiguous axial slices, contouring requires up to 140 manual delineation for each patient. This takes at least 12 min per tumor volume [21], 2 min per one OAR such as bladder and femur [22], and usually more than 2 min for OARs with complex shape such as anorectum. Furthermore, manual contouring is subject to intra- and inter-operator variability, the latter being the main source of uncertainty in treatment planning [23]. Therefore, a myriad of auto-contouring algorithms have been developed in the last decades in order to accelerate the clinical workflow and overcome the operator related variabilities [24].

Currently available auto-contouring methods can be classified into traditional methods [25] and deep learning based methods [26–28]. Traditional auto-contouring methods include model-free and knowledge-based approaches [25]. Model-free methods include thresholding and region growing [29]. They are essentially based on individual pixel intensities and are therefore susceptible to imaging artifacts, especially at the anatomical structures edge [30]. Knowledge-based approaches, such as atlas-based methods, are dependent on a priori model of the structures to be contoured [31]. The accuracy of knowledge-based approaches is largely determined by the degree of anatomical similarity between the testing subject and the atlas library subjects [32]. Thus, performance is affected by inter-subject anatomical differences, and is significantly compromised at the level of a tumor or a surgical void [32]. For example in atlas-based methods, it is not possible to create a library that is diverse enough to cover all relevant variants of human anatomy. This is more prominent in the case of cervical cancer where the tumor, enclosed in an anatomical compartment of soft tissue, has a disorganized growth pattern leading to a variety of unpredicted distorted soft tissue, vascular, and lymphatic vessels shapes.

As for deep learning, contouring is conceptually a semantic segmentation task that consists of classifying each image pixel or voxel as belonging to a predefined class. This computer vision concept has been applied in multiple fields, including landscaping, traffic surveillance, satellite imaging, and medical imaging [33]. In the latter category, deep learning based auto-contouring for radiation therapy planning has been the main focus given its potential to cost-effectively accelerate the radiation therapy workflow and improve patient care [34]. Amid different architectures, deep convolutional neural networks (DCNNs) of which the basic model was introduced by Fukushima in 1980 for pattern recognition [35], are the most reported in medical auto-segmentation with satisfying results [36]. DCNNs consist of locally connected layers, where each weighted input from a small neighborhood, known as the receptive field, is fed into the units of the next layer [36]. In contrast to the fully-connected feedforward networks where each node is in direct connection to each node in both the previous and the next layer [37], nodes in DCNNNs are rather connected to a patch of local nodes of the previous and next layer, allowing for convolution operations over a lower number of parameters. Thus, DCNNs enable the automated and fast abstraction of a large number of input features, a characteristic of particular importance in radiology where each medical image volume is composed of millions of voxels.

While traditional auto-contouring methods lack robustness and require significant manual adjustment, deep learning-based auto-contouring methods offer the advantage of being completely user-independent, easy to implement among multiple institutions, and less vulnerable to image artifacts than traditional auto-contouring methods [27,28]. As such, a myriad deep learning-based algorithms that outperformed traditional methods in automatic contouring [27,33] have been suggested in the last decade. Still, deep learning based methods are faced by the burden of procuring large labeled datasets for training and testing [38,39], which is usually difficult to achieve in the medical domain [39,40]. Under this perspective, U-Net for two-dimensional (2D) inputs was introduced in 2015 by Ronneberger et al. [41], and extended to a three-dimensional (3D) model in 2016 by Çiçek et al. [42] and achieved semantic segmentation using very few training datasets. U-Net derives its name from its "U" shaped architecture with a descending contracting pathway that condenses contextual information and a symmetric ascending expanding pathway that extracts low-level pixel information [41]. With such arrangement, U-Net architecture is able to efficiently merge high-level global to low-level local information, resulting in high performance accuracy even with few training samples. Hence, multiple articles have been published to report the segmentation outcomes with U-Net derived networks. For example, in 2017, a U-Net based fully convolutional network was used on the BRATS 2015 dataset [43] to segment brain tumors on 274 MR images [44]. Dong et al. performed five organs segmentation on 35 sets of thorax CT volumes using a combination of U-Net and generative adversarial network (GAN) [45]. Balagopal et al. used a 2D U-Net followed by a 3D U-Net model to perform multi-organ segmentation on pelvic CT images of 136 male patients [46].

In this study, we build upon the U-Net model to propose a novel approach for auto-contouring in PET/MR-based cervical cancer radiation therapy. Instead of a classic method wherein the image volume is used directly as input for the DCNN expected to directly yield the multi-class segmented output, we designed an auto-contouring strategy that mimics the human model of thinking by integrating in the pre-processing steps high level anatomical topography and radiological properties. Our

proposed tactic enables the contouring of the GTV and 4 OARs despite the limited dataset for training. To our knowledge, this article is the first to address the challenge of simultaneous GTV and OARs auto-contouring on cervical cancer PET/MR.

Compared to the previously published medical auto-segmentation papers, the novelties of our work can be summarized as follows:

(1) To decrease the training computational complexity, we employ the shallow U-Net (sU-Net) model in which the encoder depth is limited to 2. Our results show that such depth does not compromise the ability to capture high- and low-level image features and reduces the likelihood of overfitting with limited training data, as sU-Net necessitates short training and testing times while preserving a good prediction accuracy.
(2) We introduce the concept of focal and sequential training. The term focal refers to the intelligent use of FDG-PET data and anatomical topography in order to localize the structure(s) to be contoured in the center of the input. The term sequential implies a successive training of 3 consecutive sU-Nets, and a feed forward strategy in which the input of one sU-Net makes use of the preceding sU-Net output.
(3) The number of datasets used in this study is considerably smaller than the previously published reports on auto-contouring. However, the combination of a focal and sequential training along with the sU-Net structure, allow us to overcome such impediment. In fact, with only 13 datasets, our suggested approach yields an accurate automatic-contouring for the GTV and four OARs.
(4) We introduce in this study the patch dice loss (PDL) function that, in contrast to the commonly used generalized dice loss (GDL) function, accounts for the dice similarity index (DSI) at the level of each training patch. We prove that PDL results in better performance than GDL and is worth future investigation in deep-learning based automatic contouring.
(5) Finally, we apply the computational approach on cervical cancer PET/MR images as it provides better soft tissue contrast than CT, especially for the female pelvis. Given our method's remarkably short prediction time, this study represents a cornerstone for acceleration of the clinical implementation of a PET/MR based cervical cancer workflow.

2. Materials and methods

2.1. Background

2.1.1. AUTO-CONTOURING for cervical cancer radiation therapy

With the recent attentiveness for the benefits of 3D conformal therapy in cervical cancer [47], automatic contouring of the tumors and OARs became a subject of interest given its potential to accelerate the radiation therapy workflow. When compared to other anatomical sites, only few published articles have tackled the topic. Nevertheless, the clinical translation of the suggested algorithms remains faced by two main inconveniences. One, none of the methods applies to tumor and OARs simultaneously. Two, the input of these algorithms consisted of either PET, CT, or diffusion-weighted MR images which are not always available, and not of the T2-weighted, TSE-SSH MR images considered as gold standard for diagnosis and radiation therapy planning. For example, Chen et al. performed in 2019 automatic tumor segmentation on 50 sets of PET images by combining deep learning with anatomic prior knowledge [48]. The method consisted of embedding the cervical tumor geometrical shape and spatial location within a DCNN first, then relied on auto-thresholding to output the segmentation map [48]. Lin et al. employed a U-Net model for tumor contouring, but used –instead of the PET images-the diffusion-weighted MR images of 144 patients as input for training [49]. The U-Net model exhibited a triple-channel input and achieved a DSI of 0.82 on the testing set [49]. More recently, Liu et al. performed auto-contouring for OARs in cervical cancer using 105 CT volumes [50]. The DCNN model used by Liu et al. was also based on U-Net, however the encoder/decoder depth was 4, and the convolutional layers were replaced by Context Aggregation Blocks [50].

2.1.2. Loss function used for deep learning-based automatic SEGMENTATION

Many loss functions have been proposed for the task of automatic segmentation in deep learning. Some of those used early, such as cross entropy, were based on metrics more relevant for binary classification than to semantic segmentation and were tested on 2D data [51]. Most of the currently-used loss functions are based on DSI which, for 2 sets A and B, can be defined as:

$$DSI(A,B) = 2 \times |\ (A \cap B)| \ / \ (|A| + |B|) \tag{1}$$

where $|\cdot|$ indicates the cardinality (number of elements) and $\cap$ indicates the intersection [52]. DSI values are between 0 and 1, with a value of 1 indicating an exact agreement among the two contours [25]. For one 3D structure composed of N voxels, $g_{i=1,\ldots,N} \in G$ ($g_i \in \{0,1\}$) being the ground truth binary contours, and $p_{i=1,\ldots,N} \in P$ ($p_i \in [0; 1]$) being the predicted probabilistic segmentation output of the softmax layer, DSI can be expressed as follows:

$$DSI(P,G) = \frac{2\sum_{i=1}^{N} p_i g_i}{\sum_{i=1}^{N} p_i^2 + \sum_{i=1}^{N} g_i^2} \tag{2}$$

In formulating a DSI-based loss function for multi-class classification, authors needed to account for the common problem of class imbalance that derives essentially from two factors: 1) most of the voxels in a training patch belongs to the background and 2) the sizes of the contoured structures themselves are different [53]. Under this perspective, the GDL was introduced in 2017 by Sudre et al. [54] to account for highly unbalanced classes in segmentation. For K number of classes, GDL can be defined as follows [54]:

$$GDL(P,G) = 1 - \frac{2\sum_{k=1}^{K} w_k \sum_{i=1}^{N} p_{ki} g_{ki}}{\sum_{k=1}^{K} w_k \sum_{i=1}^{N} (p_{ki} + g_{ki})} \tag{3}$$

where w_k is a weighting factor that scales each class by the inverse of its size using the below equation [54]:

$$w_k = \frac{1}{\left(\sum_{i=1}^{N} g_{ki}\right)^2} \tag{4}$$

2.2. The proposed workflow

2.2.1. Network architecture

The proposed sU-Net is a 3D U-Net that consists of a contracting encoder path followed by an expanding decoder path [42]. The structure of the sU-Net model is shown in Fig. 1. The contracting path includes a convolutional layer [55], followed by batch normalization [56], rectified linear unit [57], and a $2 \times 2 \times 2$ max pooling with a strides of 2 in each dimension [58]. With this arrangement, features resulting from an input image filtered through the convolutional layer kernels will be regularized via the batch normalization layer by their sample mean and standard deviation (SD) [56]. The batch normalization layer will significantly accelerate the sU-Net training, and the following rectified linear unit layer will further speed-up the convergence of the gradient descent [59]. Max pooling is then used to pass the maximum filter response of the local region to the next layer [60]. Max pooling layers aggregate the features of a convolutional layer, thus decreasing the number of parameters, which leads to faster convergence and prevents overfitting [61]. The expanding

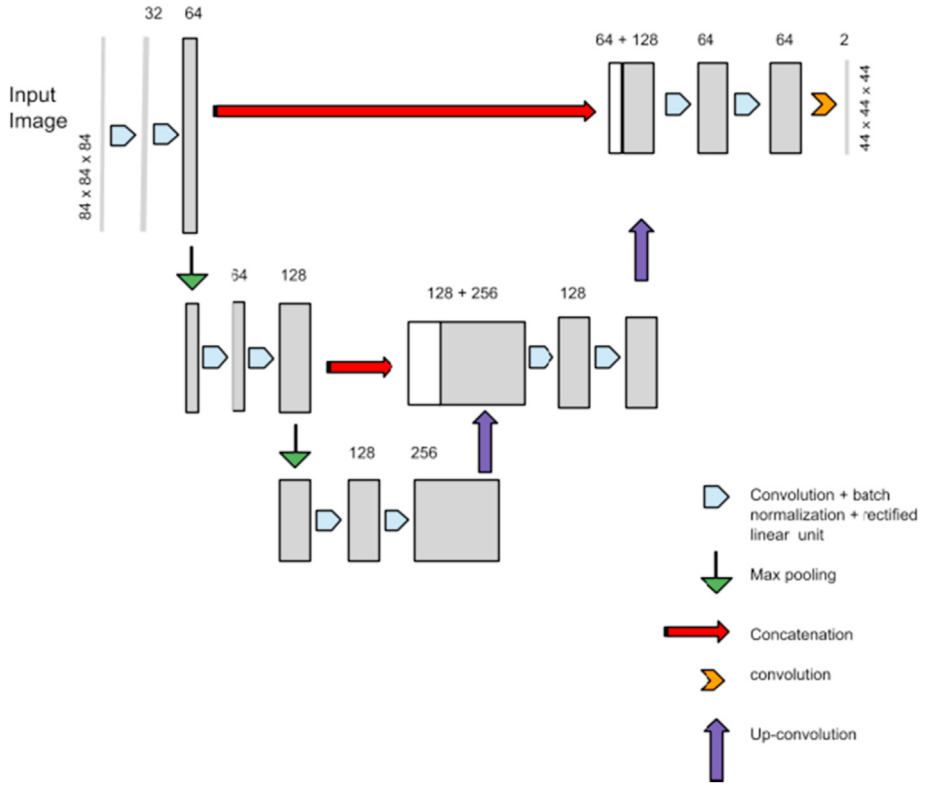


Fig. 1. sU-Net architecture for a one channel input and a binary classification. The number of features appears on top of the box. The white boxes correspond to the concatenated feature maps.

path is used to recover the initial image resolution by making use of up-convolution that entails an up-sampling procedure to double the input features dimension followed by convolution [62]. Finally, two concatenation connections are inserted between the descending and the ascending pathways to ensure a multi-scale features integration and improve segmentation [63]. The last layer of sU-Net is a softmax layer that normalizes the input of the previous convolutional layer into a probabilistic distribution for each segmentation class [64].

Compared to the previously published reports with U-Nets [41], two main modifications are employed in our study. First, we limited the encoder depth to 2 instead of the usual 3 [41,42,44]. Second, the sU-Net input size was set to 84 × 84 × 84 voxels in x,y, and z directions respectively, making the last layer result in an output size of 44 × 44 × 44 voxels due to the use of convolution padding. With these modifications, sU-Net model accommodates the task of multi-class segmentation in the settings of limited data availability, by decreasing the total network size and the subsequent number of parameters to be optimized. With a voxel size of 3.2 mm × 3.2 mm × 5 mm, the receptive field of each output voxel is of 128 mm × 128 mm × 200 mm, preserving the sU-Net capability of broad context, high-level learning.

2.2.2. Patch dice loss function

Loss functions should account for class imbalance between background and structure of interest, or among structures of interest themselves in case of multi-class classification. In addition, loss functions should be customized to the anatomical sites and training scheme used in the methodology. More specifically for our method where training is achieved using random patch extraction for five structures of interest, it is inevitable that many of patches will not include all the structures, posing the issue of negative samples during training [65]. Furthermore, for a training scheme that combines, at each iteration, image patches derived from different image input, this GDL formulation does not directly account for the DSI of each image patch as it lumps all the voxels of one iteration into one whole set of N pixels. Therefore, we introduce in this manuscript the PDL function. Assuming a mini-batch size of M, where each input of the mini-batch is a 3D patch indexed by (*x,y,z*), the DSI for a class *k* and a patch *m* can be written as follows:

$$DSI_{km}(P,G) = \frac{2\sum_x\sum_y\sum_z g_{xyzkm}p_{xyzkm}}{\sum_x\sum_y\sum_z g^2_{xyzkm} + \sum_x\sum_y\sum_z p^2_{xyzkm}} \quad (5)$$

Using the above formulation of DSI_{km} the PDL is defined as the weighted average over classes and unweighted average over the mini-batch by the following equation:

$$PDL(P,G) = 1 - \frac{1}{M}\sum_{k=1}^{K}\sum_{m=1}^{M} W_k DSI_{km}(P,G) \quad (6)$$

With this formula, PDL computes the score function at each iteration as the arithmetic mean of all the patches' DSI. This is different from GDL whose score reflects a weighted lump of all the combined pixels, without considering the DSI at each individual patch. As all classes – including background – are weighted equally ($W_k = 1/K$), this PDL formulation eliminated the influence of region size discrepancies. For a specified pixel (x = a, y = b, z = c, k = d, m = e), PDL can be differentiated for the prediction $p_{abcde\ [fx]}$ using the quotient rule for derivatives, yielding the gradient:

$$\frac{\partial PDL}{\partial p_{abcde}} = \frac{2W_k}{M}\left(\frac{2\left(\sum_x\sum_y\sum_z g_{xyzde}p_{xyzde}\right)p_{abcde} - g_{abcde}\left(\sum_x\sum_y\sum_z g^2_{xyzde} + \sum_x\sum_y\sum_z p^2_{xyzde}\right)}{\left(\sum_x\sum_y\sum_z g^2_{xyzde} + \sum_x\sum_y\sum_z p^2_{xyzde}\right)^2}\right) \quad (7)$$

2.2.3. Focal and sequential approach

The proposed approach is inspired by the human model of thinking. In fact, the human cognition is grossly divided into three concepts: granulation, organization, and causation [25]. We can therefore postulate that the human brain begins with granulation to decompose a given image into different sub-volumes with each sub-volume containing an anatomical structure. Thereafter, granulation is followed by organization to classify each pixel in the sub-volume into structure or background [66].

In the setting of cervical cancer, the brain will use knowledge of the general female pelvis topography to the PET/MR features in order to virtually decompose the image. While individual pelvic organs present inter-patient variability, the general anatomical topography is well preserved. This topography is represented in Fig. 2, and can be summarized as follows: the bladder is always situated in the center of the pelvic cavity, and it is anterior to the female reproductive system from where the GTV arises. The female pelvic skeleton displays a left/right symmetry of the femurs in respect to the bladder. The anorectum is in the posterior middle part of the pelvis, and posterior to the anterior part of the femurs. Knowing the radiological properties and general organization of the female pelvis anatomical landmarks, the human brain will relate to each structure, a specific sub-volume. For example, the radiation oncologist will expect to find the bladder and GTV in an area of high SUV, will not contour the right femur on the left side of the pelvis, and will not perceive the rectum to be located more lateral to the right or left femurs.

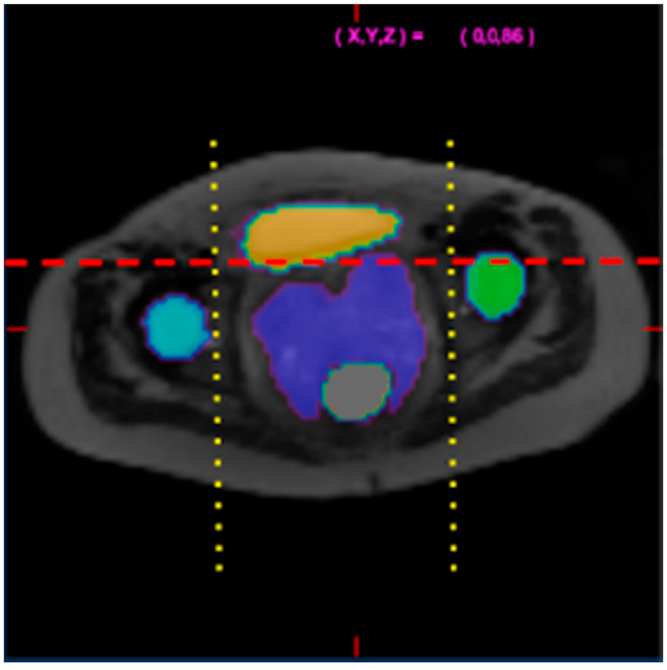


Fig. 2. Axial view of a female pelvis in a patient with cervical cancer. The bladder (orange) is located in the central part of the pelvis. Right femur (blue) and left femur (green) are located symmetrically from each side of the bladder (dotted yellow lines). The bladder is situated anteriorly to the cervix cancer tissue (violet), and the rectum (grey) is located posteriorly to the anterior surface of the femurs (dotted red line). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

In order to translate the granulation brain processes to the computer we designed a focal and sequential approach for auto-contouring (Fig. 3). An intelligent images processing is proposed to simulate the granulation of the human brain. In the image processing steps, there is no categorization of pixels into a specific class, but we rather use the prior knowledge of radiological features and anatomical topography to grossly determine the sub-volume where a structure of interest, and then process this sub-volume to be centered in the next step of sU-Net training. At last, for organization or categorization, we are taking advantage of the sU-Net architecture to leverage high- and low-level features, and classify each pixel into a specific category. Similar to granulation, the algorithms uses the PET/MR properties of the bladder and the GTV explained in the introduction, and the above-noted general anatomical topography, to decompose the PET/MR image volume into four sub-volumes containing respectively: GTV and Bladder in the first step, left and right femurs in the second step, and anorectum in the third step. The organization is simulated by using the focal sU-Net training scheme, wherein the structures of

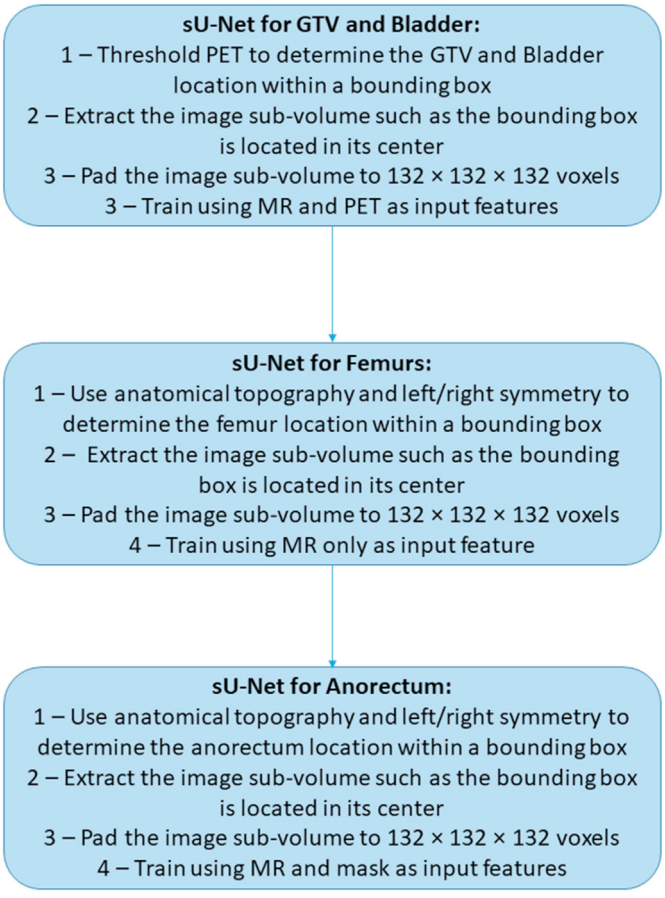


Fig. 3. The general workflow for sU-Net sequential and focal training in PET/MR based cervical cancer.

interest are placed in the center of the sub-volumes, and the output is organized to mainly enclose the centrally located voxels, and ignore the peripherally located voxels.

2.2.4. Image volumes propressing

In this section, we describe the image processing steps used to decompose each initial image volume into the sub-volumes that will be used for the sU-Net training.

(A) GTV AND BLADDER AUTO-CONTOURING

FDG-PET images are first thresholded to above 0.75 of the normalized intensity. After thresholding, the bounding box containing the two central components of the thresholded images that correspond presumably to the GTV and bladder is determined. PET and MR image sub-volumes are then extracted so that the bounding box is located in the center of the sub-volume. A two-class segmentation training is performed using the PET and MR images as input features.

(B) BILATERAL FEMURS AUTO-CONTOURING

Using the female pelvic skeleton anatomical left/right symmetry, we can train one sU-Net for contouring both femurs. For the left femur, the area extending from the bladder bounding box left lateral surface leftwards is considered as containing the left femur. Similarly for the right femur, the area extending from the bladder bounding box right lateral surface rightwards is considered as containing the right femur. Separate MR image sub-volumes for each femur are extracted such as the bounding box of the femur is located in the center of its corresponding sub-volume.

(C) ANOCRECTUM AUTO-CONTOURING

As the anorectum is consistently located in the posterior pelvis, the anterior surface of the predicted femur contours in the second training is chosen as the anterior boundary of the anorectum bounding box. The left and right boundaries are considered to be the same as the bladder. In addition to the zero padded MR images, we also include in the input a mask where we masked the voxels located outside the bounding box and those belonging to the GTV, bladder, and femurs.

2.3. Experiment

2.3.1. Data acquisiton and CONTOURING

A retrospective, IRB-approved study was conducted at University Hospitals Cleveland Medical Center to include adult female patients treated for a biopsy-proven cervical cancer between June 2015 and June 2018. Among these patients, 13 had undergone a PET/MR using a Philips Ingenuity TF PET/MR system [67,68] per our institution protocol: Field of view of 300 mm, echo time (TE) of 80 ms, repetition time(TR) of 1097 ms, and a total scan duration of 60 s. Patients' age, radiological International Federation of Gynecology and Obstetrics (FIGO) stages, TNM stages, and image voxel sizes and spacing are summarized in Table 1.

MR and FDG-PET images were de-identified and uploaded into MIM (MIM software, Cleveland, OH). T2 weighted, TSE-SSH MR images were then fused with the PET images to contour the GTV and bladder. Right and left femurs and anorectum contours were delineated using only the TSE-SSH MR T2-weighted MR images. Manual contouring was performed in MIM by a junior radiation oncology resident (AB), then reviewed and approved by an attending radiation oncologist with expertise in gynecologic malignancies (BJT). OARs were delineated according to the guidelines defined by the radiation therapy oncology group consensus panel [14]. Images and contours were re-sampled in MIM to a common pixel spacing of $3.2 \times 3.2 \times 5\ mm^3$. An example for the TSE-SSH MR and PET images is featured in Fig. 4. PET and T2-weighted, TSE-SSH MR images were then loaded into MATLAB 2019b (MathWorks, Inc.) using COMKAT Image Tool [69,70], and their intensities were normalized to [0, 1] for further processing and training.

2.3.2. Setup

The re-sampled PET/MR image volumes were processed as described in paragraph III.D in order to generate the four sub-volumes. Each 3D image sub-volume and its ground truth contours is then zero padded to a common size of $132 \times 132 \times 132$ voxels. From these image sub-volumes, 3D patches sized $84 \times 84 \times 84$ voxels each, were extracted to serve as input for the 3D sU-Net. As no padding was used during training, the output size was smaller than the input. As such, the predicted sU-Net output area was limited to the central area of the input patch, and the output size was $44 \times 44 \times 44$ voxels. Each of the three steps was performed using a leave-one-out strategy, where 12 (first and third training) or 24 (second training) image sub-volumes are used as training set, and the remaining image sub-volume(s) is (are) used for testing. In the first and third training steps, the mini-batch size was set to 8, the number of patches per image sub-volume was 24, and the number of epochs was 50. For the femurs' training step, two image sub-volumes can be extracted from each original MR image volume, one sub-volume for each femur. Thus, the total number of data sets used in the femurs experience was 24. The number of patches per image sub-volume was 16, the mini-batch size was 6, and the number of epochs was 40. Right/left flipping was used for data augmentation. After initialization of the DCNNs weights, the back-propagation technique, wherein the loss function is efficiently minimized or maximized using its calculated derivative, is used for weights optimization. An Adam random gradient descent method was applied [71]. . An example of PET thresholding, central components isolation, and images used as input features for each training step are featured in Fig. 5.

Table 1
Age, Histology, FIGO radiological staging, PET/MR voxels spacing and size for the 13 patients included in our study.

Subject Number	Age at Diagnosis (Y)	Histology	FIGO Radiological Staging	PET/MR Voxel Spacing (mm × mm × mm)	PET/MR Size (voxels)	PET/MR Size after Re-sampling (voxels)
1	79	SCC	IIIB	0.62 × 0.62 × 5	672 × 672 × 50	132 × 132 × 50
2	38	SCC	IIIA	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
3	81	SCC	IIB	0.63 × 0.63 × 5	528 × 528 × 50	104 × 104 × 50
4	70	SCC	IVB	0.67 × 0.67 × 5	640 × 640 × 50	134 × 134 × 50
5	79	Adeno	IIB	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
6	57	SCC	IIIB	0.63 × 0.63 × 5	528 × 528 × 50	104 × 104 × 50
7	82	SCC	IIB	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
8	62	SCC	IIB	0.54 × 0.54 × 4	576 × 576 × 55	97 × 97 × 44
9	49	SCC	IIB	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
10	37	SCC	IIB	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
11	52	Mixed	IIA	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
12	45	SCC	IIB	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44
13	83	Adeno	IB	0.53 × 0.53 × 4	672 × 672 × 55	112 × 112 × 44

SCC = Squamous Cell Carcinoma, Adeno = Adenocarcinoma, Mixed = Adeno/SCC.

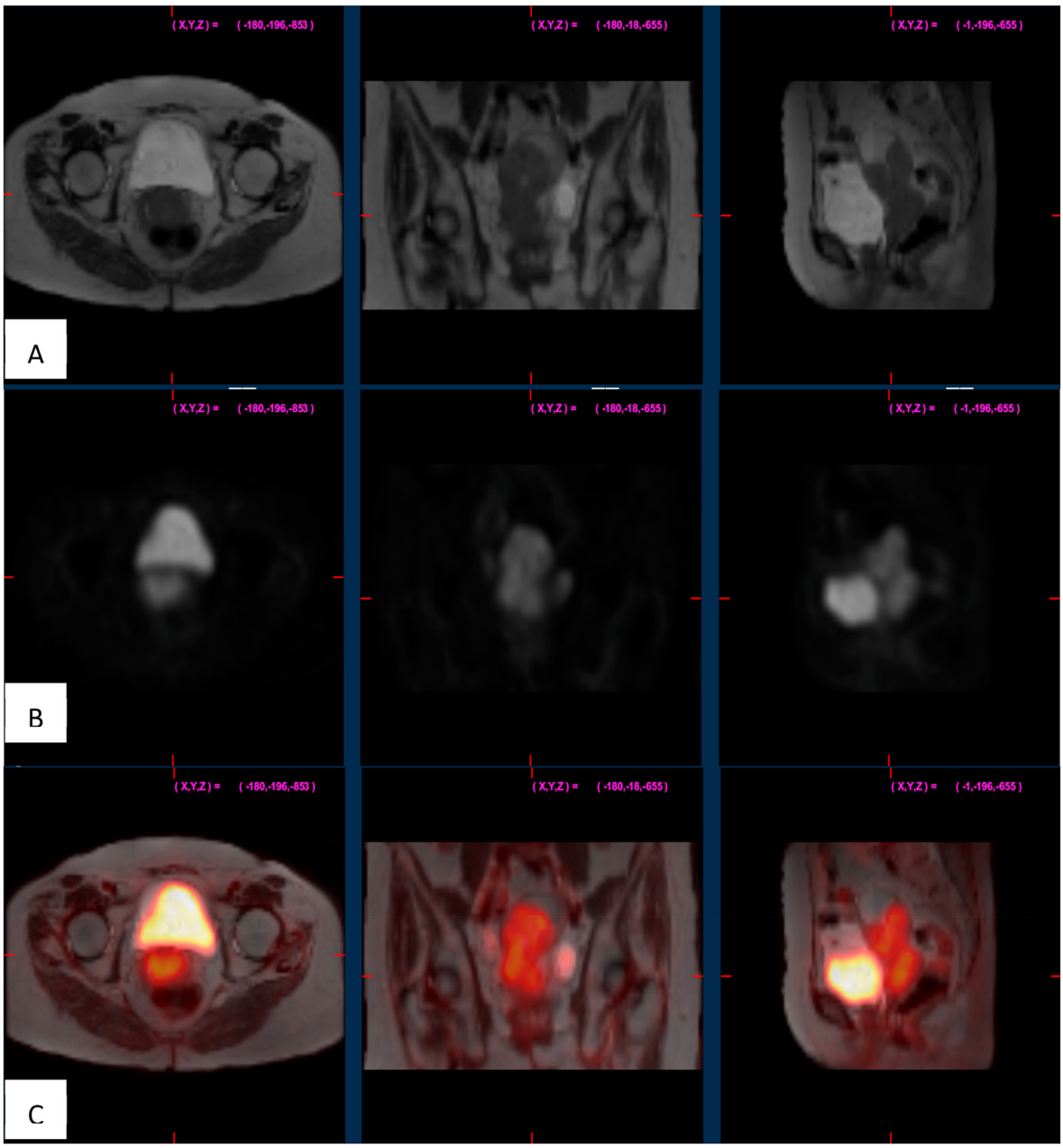


Fig. 4. Axial, coronal, and sagittal (left to right) views of TSE-SSH MR (A), PET (B), and fused MR/PET (C) images.

In addition to the focal and sequential sU-Net approach, we also performed the training experiment a classic approach using U-Net-with an encoder/decoder depth of 3. The workflow is similar to the previously published U-Net papers [42,62] and consisted of using the normalized images as input and the 6 classes segmentation map (background, GTV, bladder, right femur, left femur, anorectum) as output. While there is no currently gold standard method for GTV and OARs auto-contouring for FDG-PET/MR cervical cancer images, we believe that the classic U-Net approach serves as a satisfactory benchmark comparative method to evaluate both, the PDL and our proposed approach. In order to optimize sampling and increase the training samples, the patch size was set to 64 $\times$ 64 $\times$ 64 voxels after zero padding the image volumes 132 $\times$ 132 $\times$ 132. The experimental setup for sU-Net and U-Net is summarized in Table 2.

2.3.3. Labels prediction

During testing, the image volume is first decomposed into its sub-volumes as described in paragraph III.D. For sU-Net, the test image sub-volume is then zero padded to a total size of 172 $\times$ 172 $\times$ 172 voxels. Afterwards, an overlap-tile strategy is used [41] where labels are predicted using 84 $\times$ 84 $\times$ 84 voxel-sized patches, then recombined to give the prediction over the zero padded image sub-volume. Finally, the image sub-volumes are cropped to their original size and subsequently combined to reconstitute the original image volume. A similar strategy was adapted for the U-Net prediction step.

2.3.4. Performance evaluation

Training and prediction times were recorded. In addition to visual inspection, automatically-defined contours are scored using three indices: dice similarity index (DSI) [52], the Jaccard similarity index (JSI) [72], and the boundary F1 score (BF) [73].

DSI was defined earlier in this manuscript as it is used in the loss function. As for JSI, it quantifies the overlap between the ground truth and the predicted labels as the ratio of the intersection over union:

$$JSI(A, B) = |(A \cap B)|/|(A \cup B)| \quad (8)$$

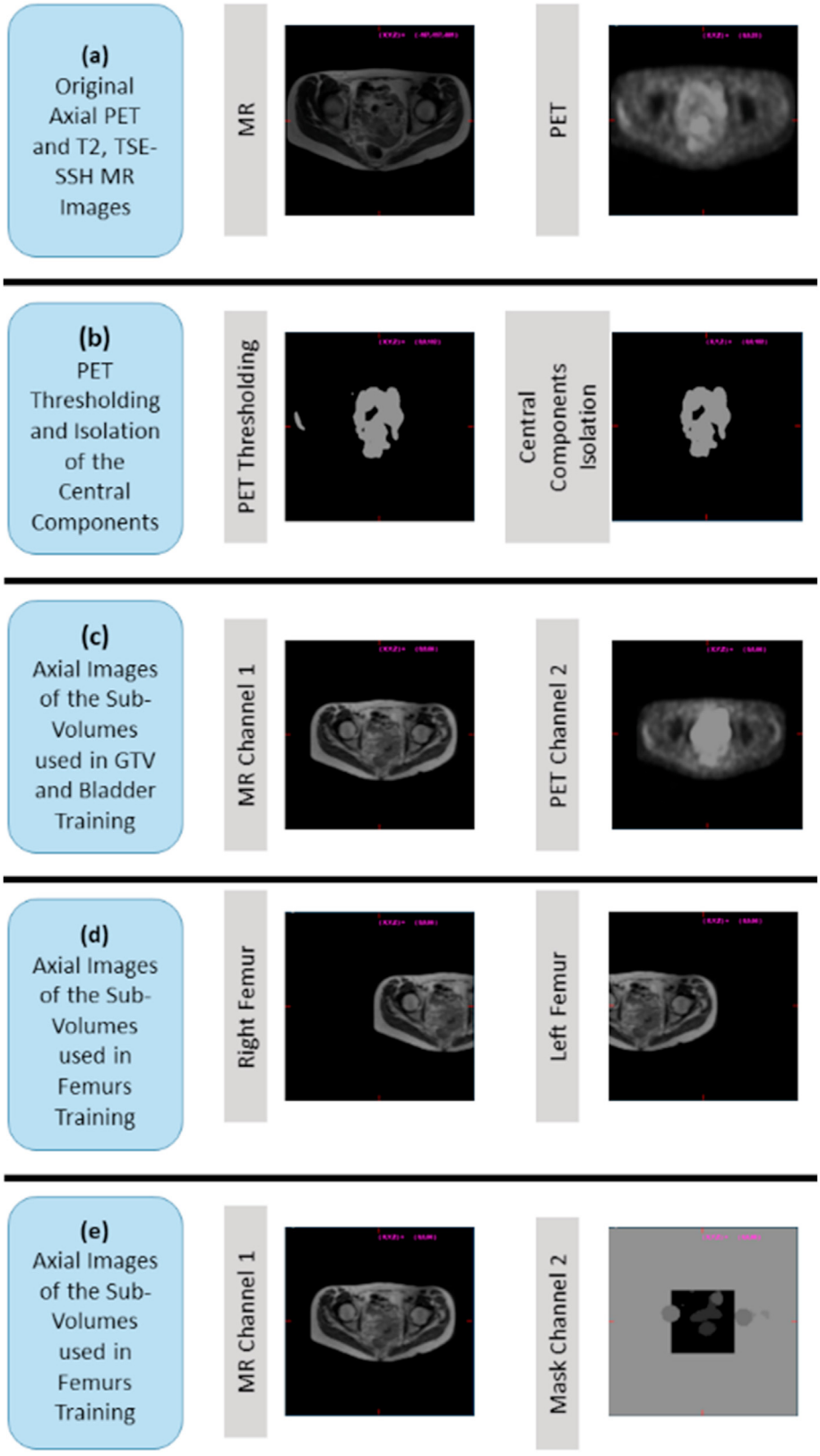


Fig. 5. Example of image processing and training inputs for one subject.

where ∪ represents the union. JSI calculated values are between 0 and 1, with 1 corresponding to perfect matching between ground truth and prediction. JSI accounts simultaneously for correctly and incorrectly labeled pixels [73]. It is related to DSI by the following equation:

$$JSI(A,B) = |DSI(A,B)|/|2 - DSI(A,B)| \quad (9)$$

Both DSI and JSI are focused on the prediction accuracy, but do not address the segmentation boundary itself [73]. For this, we used the BF as a complementary index for JSI and DSI to reflect the overlap of the

Table 2
Summary of the experimental setup for sU-Net and U-Net.

	sU-Net Focal and Sequential Approach			U-Net Classic Approach
	GTV and Bladder Auto-Contouring	Femurs Auto-Contouring	Anorectum Auto-Contouring	Multi-Class Classic Approach
Number of Leave-one-out Repeats	13	13	13	13
Number of 3D Image Sub-volumes Used for Training in Each Repeat	12	24	12	12
Number of 3D Image Sub-volumes Used for Prediction in Each Repeat	1	2	1	1
Number of Training Patches Extracted from Each 3D Image Sub-volumes in Each Repeat	24	16	24	32
Total Number of Training Patches in Each Repeat	288	384	288	384
Mini-batch Size in Each Repeat	8	6	8	8
Number of Epochs in Each Repeat	50	40	50	50
Segmentation Classes	1 – Background 2 – GTV 3 - Bladder	1 – Background 2 - Femur	1 – Background 2 - Anorectum	1 – Background 2 – GTV 3 – Bladder 4 – Right Femur 5 – Left Femur 6 - Anorectum

contour boundary [73]. For one class binary classifier, BF is defined as follows:

$$BF = 2 \times precision \times recall/(recall + precision) \tag{10}$$

where precision or positive predictive value is defined as [74]:

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

while TP refers to the number of true positive voxels defined as the number of voxels that are correctly as belonging to the class of interest, and FP refers to the number of false positive voxels defined as the number of background voxels identified by the classifier as belonging to the class of interest. FN refers to the number of false negative defined as the number of class of interest voxels falsely identified by the classifier as belonging to the background. Recall or sensitivity is defined as [74]:

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

For multi-class classification, BF is averaged over all classes so it results in one value for the whole image set, with a value of 1 corresponding to perfect matching [73].

Training and prediction times, DSI, JSI, and BF scores at each step were compared using a two-tailed paired *t*-test, with a significance level of 5%. In order to avoid correction for multiple comparison, and since no further clinically relevant information would be concluded, the overall metrics and the total prediction and training times were not statistically compared.

3. Results

Fig. 6 shows a T2 TSE-SSH MR image with manually- and automatically-defined contours of the GTV and the four OARs. Fig. 7 summarizes, using bar graphs, the segmentation results, the training, and prediction times for the two training schemes using PDL and GDL. Contours of two best achieving methods, i.e. sU-Net PDL and sU-Net GDL, are directly compared on an axial slice in Fig. 8. By visual inspection, the deep learning based contours with sU-Net are similar to the ground truth manual contours. The classic U-Net contours with GDL exhibit a significant deviation from the ground truth, show no smooth boundaries, and tend to preferably label pixels as bladder. With PDL, the classic U-Net contours feature an overall smooth arrangement, with good high-level localization of the central structures such as GTV and bladder, but with a clear decreased accuracy compared to the contours obtained with the sU-Net focal and sequential approach. Summary statistics for DSI, JSI, BF, training and prediction times are given in Tables 3–5.

The classic methodology with U-Net achieved a satisfactory DSI only for the bladder, when PDL was used as the loss function. Among the five structures, classic U-Net achieved better performance for GTV and bladder, compared to femurs and anorectum. In the classic U-Net training scheme, PDL was clearly superior to GDL in terms of DSI, JSI, and BF. GDL was slightly more rapid than GDL in the classic U-Net training. Nevertheless the 3 min advantage in the training time, and the 0.42 s advantage in the prediction time provide no clinical advantage, as training time is done offline, and 0.42 s is negligible in the current clinical settings. Compared the classic U-Net, our proposed methodology showed a clear superiority as it achieved a DSI above 0.7 for 4 structures and overall, with only the anorectum accomplishing a DSI of 0.6.

As for sU-Net, PDL performed better than GDL for all structures, except GTV. Statistically, the difference between the performance measure was only significant at the recorded times. Except for the training step for left and right femur contouring and the prediction step for left femur contouring, PDL always exhibited shorter times than GDL. The overall training time was also shorter for PDL than GDL, and ranged between 62.55 and 133.4 min per sU-Net training for PDL. Prediction times were extremely short with 1.9 s per contour for PDL and 2.18 s per contour for GDL. Numerically, the maximum DSI difference between PDL and GDL was 0.1 (GTV, bladder, and femurs), and the minimum DSI difference was 0.02 for the anorectum. As for the overall DSI, it was higher using PDL than GDL (0.78 vs. 0.77), with GTV and 3 out of the 4 OARs reaching a DSI above 0.7. DSI was the highest for solid organs (right and left femur), and the lowest for soft, air-filled organs (anorectum). JSI ranged between 0.45 and 0.79 for PDL, and 0.43 to 0.77 for GDL. For each contour, JSI values were lower than DSI values. Similarly to DSI, JSI was the highest for femurs and the lowest for anorectum, and JSI was higher with PDL than GDL except for the GTV. As for JSI difference between PDL and GDL, it ranged between 0.2 (bladder and right femur) and 0.02 (anorectum). The predicted contour boundaries were smooth overall, having no spurious appendages as evaluated by visual inspection. Moreover, the predicted boundaries matched the ground truth boundaries as manifested in the BF indices reaching an overall value of 0.86 for GDL and 0.87 for PGL. BF values were greater or equal to 0.9 for the bladder and the femurs, approximately 0.8 for the GTV, and the lowest for the anorectum with a BF score of 0.68 with GDL and 0.71 with PDL.

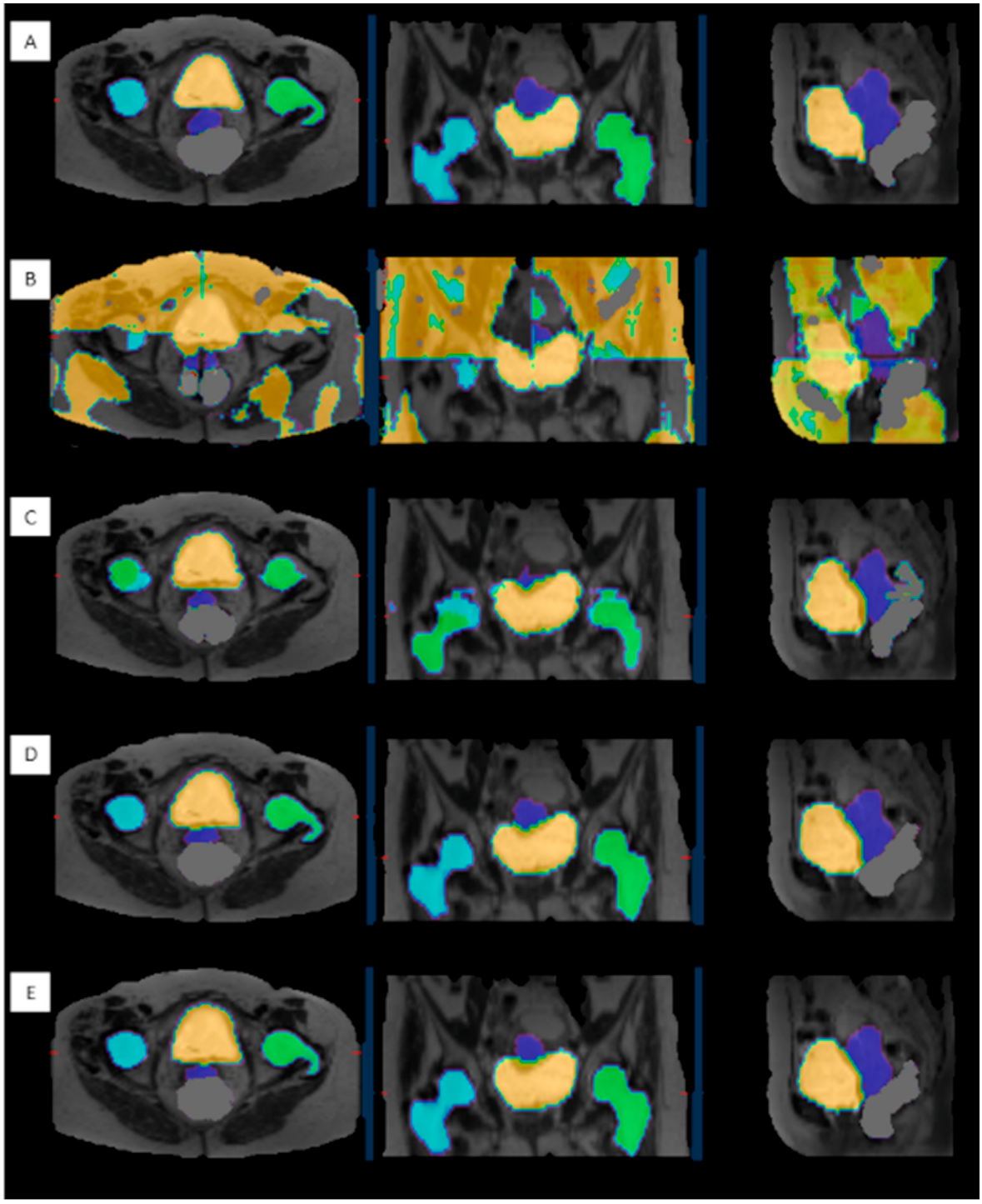
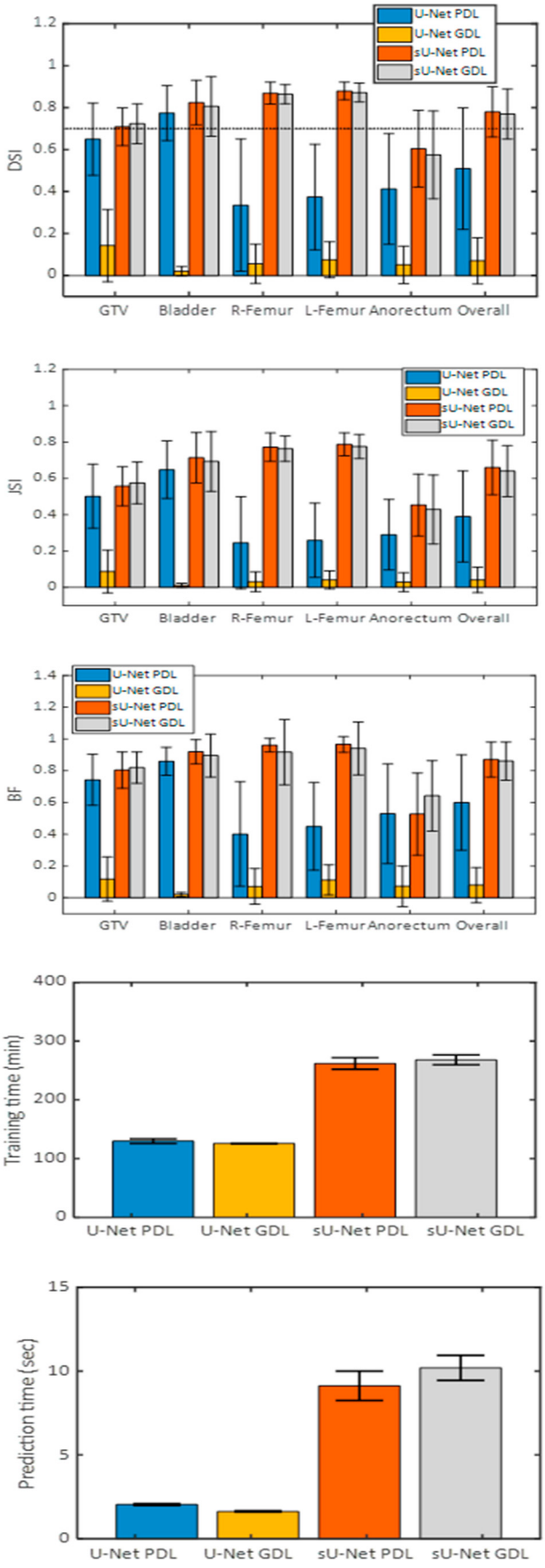


Fig. 6. Axial, coronal, and sagittal (left to right) views of the ground truth manual contours (A), autocontours using classic U-Net GDL (B), auto-contours using classic U-Net PDL (C), autocontours using focal, sequential sU-Net GDL (D), auto-contours using focal, sequential U-Net PDL (E) GTV: Violet, Right Femur: blue, Left Femur: Green, Bladder: Orange, Anoerctum: Grey. Among the contours generated with the classic U-Net approach, those obtained with PDL exhibit a better anatomical localization with the ground truth contours than those obtained with GDL. Nonetheless, the accuracy of the classic U-Net approach with PDL seems to be visually less than the accuracy of the contours obtained with the focal, sequential sU-Net. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

Deep learning is granting the medical community a transformative tool with the potential to accelerate clinical workflows, reduce health expenses, and improve the overall outcomes. The main hindrance for incorporation of deep learning methods in the daily clinical activities is the limited availability of very large sets of labeled medical data for training, validation, and testing. Particularly in the field of medical imaging, such hindrance is more pronounced given the enormous burden of acquiring consistent image sequences among different patients and institutions, in addition to the expensive expertise time required for accurate labeling. With its architecture of a contracting and expanding



(caption on next column)

Fig. 7. Bar graphs summarizing the segmentation results, the training, and prediction times for the classic U-Net and the focal, sequential sU-Net schemes using PDL and GDL. In the DSI graph, the common threshold for clinical acceptability (DSI = 0.7) is represented as a dashed horizontal line. Note that the maximum achievable value for DSI, JSI, and BF is 1. However, the corresponding y-axis for these bar graphs has been extended beyond 1 for the purpose of graphical illustration, and to prevent truncation of the error bars. As depicted visually, sU-Net achieves a better performance overall and for each structure.

pathway joined by concatenation layers, U-Net reduced the number of datasets needed for training, and accurate automatic contours had been achieved with 35 datasets [45]. Of course, such algorithms require a customizable pre-processing to adapt the network architecture to the anatomical region, data size, and number of classes. For example, a direct "plug and play" method in the case of multi-class auto-contouring for cervical cancer would have been inappropriate for the peripherally located OARs such as the femurs and anorectum, especially with the use of convolution padding. Instead, we based the pre-processing in our experiment on the model of human thinking. In fact, the human cognition is grossly divided into three concepts: granulation, organization, and causing [25]. We can therefore postulate that the human brain begins with granulation to decompose a given image into different patches with each patch containing an anatomical structure. Granulation is followed by organization to classify each pixel in the patch into structure or background [66]. Using anatomical topography and FDG-PET imaging properties, we introduced in this study the concept of focal and sequential training that imitates the human model of granulation, and overcome simultaneously the barriers of data availability and the possible neglect of peripherally located structures with the use of convolution padding. Compared to the classic U-Net, the sU-Net model reduced further number of weights to be optimized, rendering the auto-contouring task for GTV and 4 OARs overall feasible using only 13 datasets. While operations such as rotation, left/right flipping, and up/down turning have been suggested for data augmentation, we only used left/right flipping in this experiment as it was the only operation that preserved the overall pelvic anatomic topography. Finally, the suggested PDL presents a balanced function that is robust to the problem of intra- and inter-patients class imbalance. It integrates the DSI performance at the level of each patch and yields slightly better metrics than GDL for all the OARs. As for the GTV, the DSI difference was 0.1 higher for GDL than PDL.

With a sample mean contouring time of 1.9 s per contour for PDL and 2.8 s per contour for GDL, our proposed algorithm significantly reduces the radiation therapy planning time when compared to manual contouring that usually consumes more than 1 h of the physician time [20]. A DSI value above 0.7 is a validated surrogate for good overlap [75]. Therefore, our results are overall satisfactory, especially that deep learning based auto-segmentation decreases intra- and inter-observer variability commonly seen with manual contouring [34]. Specifically for the GTV, the obtained DSI values, 0.72 with GDL and 0.71 with PDL, are similar to DSI values reported when comparing manually contoured GTV using PET versus MRI. For example, Zhang et al. obtained a DSI ranging between 0.64 and 0.68 when they compared GTVs manually-contoured using MR vs. PET [16]. In our method, the range of JSI (0.34 between anorectum and right femur) and DSI (0.29 between anorectum and right femur) is not related to a low precision of our method, but rather to differences in shape, contrast, and consistency between anorectum and femur. In fact, this is an expected behavior seen similarly with human radiation oncologist whose contours are more overlapping in solid structures such as femur than soft tissue structure such as anorectum [23]. The low anorectum DSI is related to the organ properties itself given its soft tissue consistency and variable shape that depends on the bladder position and the stool and air load in the colon. Thus, manually drawn contours for the anorectal region demonstrate limited consistency even among radiation oncologists. For example, Wang et al. conducted a study using T2-Weighted MR for 93 patients with

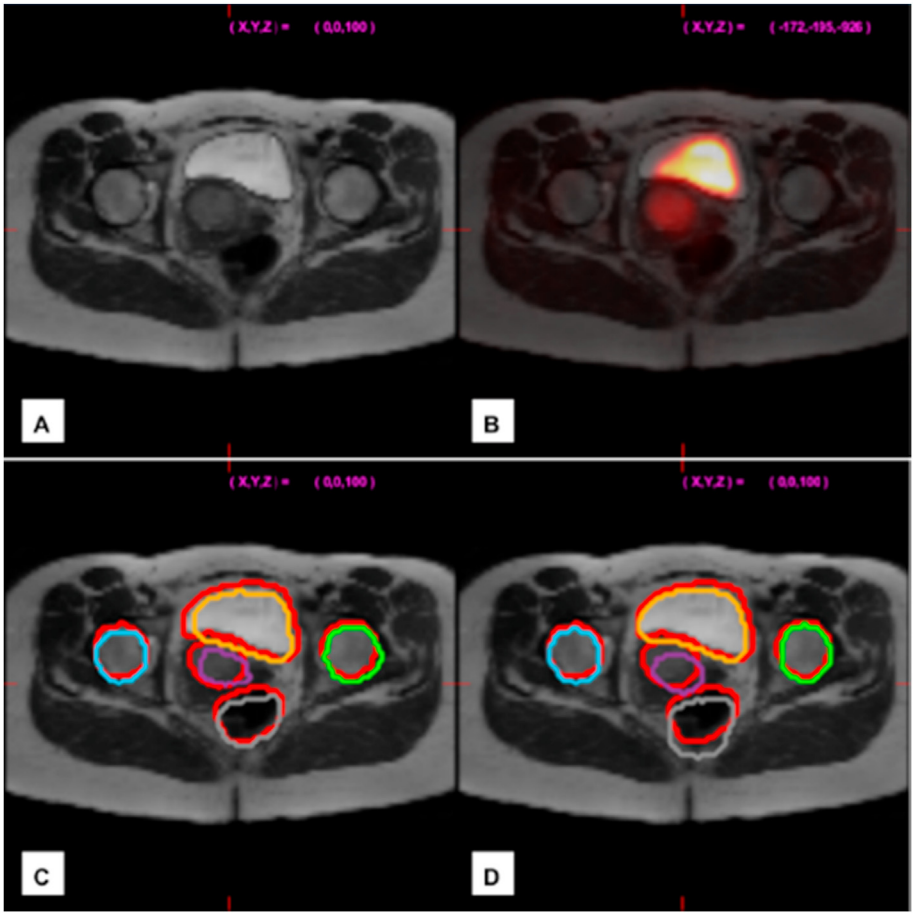


Fig. 8. Axial pelvic MR (A) and Fused PET/MR (B) for a patient with cervical cancer, and auto-contours results obtained with focal, sequential sU-Net using PDL (C) and GDL (D). Ground truth manual contours are highlighted in red. Autocontours GTV: Violet, Right Femur: blue, Left Femur: Green, Bladder: Orange, Anoerctum: Grey. Note the significant visual overlap for all the structures, except for the anorectum. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Sample mean ± standard deviation and p-values for DSI, JSI, BF, training and prediction times for the Classic U-Net.

Loss Function	GTV		Bladder		Right Femur		Left Femur		Anorectum		Overall	
	GDL	PDL	GDL	PDL	GDL	PDL	GDL	PDL	GDL	PDL	GDL	PDL
DSI	0.14 ±0.17	**0.65 ±0.17**	0.02 ±0.02	**0.77 ±0.13**	0.05 ±0.09	**0.34 ±0.32**	0.08 ±0.09	**0.37 ±0.25**	0.05 ±0.09	**0.41 ±0.26**	0.07 ±0.11	**0.51 ±0.29**
p-value	**3.99×10^{-7}**		**5.89×10^{-11}**		**5.30×10^{-3}**		**6.09×10^{-4}**		**8.54×10^{-4}**		**1.98×10^{-18}**	
JSI	0.08 ±0.12	**0.50 ±0.18**	0.01 ±0.01	**0.65 ±0.16**	0.03 ±0.05	**0.25 ±0.25**	0.04 ±0.05	**0.26 ±0.20**	0.03 ±0.05	**0.29 ±0.19**	0.04 ±0.07	**0.39 ±0.25**
p-value	**9.15×10^{-7}**		**4.52×10^{-9}**		**6.70×10^{-3}**		**1.40×10^{-3}**		**7.44×10^{-4}**		**5.71×10^{-17}**	
BF	0.12 ±0.14	**0.74 ±0.16**	0.02 ±0.01	**0.86 ±0.09**	0.07 ±0.11	**0.40 ±0.33**	0.11 ±0.09	**0.45 ±0.28**	0.07 ±0.13	**0.53 ±0.31**	0.08 ±0.11	**0.60 ±0.30**
p-value	**6.21×10^{-9}**		**5.49×10^{-13}**		**2.10×10^{-3}**		**4.96×10^{-4}**		**1.06×10^{-3}**		**8.47×10^{-20}**	

The best value achieved (highest for DSI, JSI, and BF) are indicated in bold. P-Values <0.05 are also indicated in bold.

rectal cancer and reported that the DSI between two human radiation oncologists was 0.71 [76].

The incorporation of a priori knowledge in semantic segmentation has been always used to optimize contouring results. The proposed human-inspired focal and sequential approach inherently possesses the advantage of directing the attention of the network to the structure of interest, thus bypassing the need of incorporating other strategies such as attention layers or additional decoding/encoding steps. At a computational level, the image processing steps of the proposed algorithms are mainly derived from the human intelligence model, and this seems to be the main advantage of our proposed methodology over the classic U-Net method. In fact, our incorporated a priori knowledge is ubiquitous, based on anatomical landmarks depicted regardless of the imaging quality, and not linked to a specific atlas or library. Moreover, we normalized FDG PET SUV intensity for each patient to [0; 1], so our computational workflow is consistent among different patients. Knowing that the bladder and GTV usually exhibit the highest PET SUV, we therefore thresholded the images in respect to normalized intensity, not the absolute intensity. All these advantages make our algorithm easily reproducible among institutions and researchers, especially that it is appended with a simplified DCNN structure.

PDL has clearly bypassed GDL in the classic U-Net training. Nonetheless, PDL has not outperformed GDL in a statistically significant matter with sU-Net. The clear over-performance of PDL over GDL in the classic U-Net training can be related to particularly difficult task of FDG-PET/MR auto-contouring: the input consist of two multi-modality images, the task implies a multi-class classification into six categories, and the absence of intelligent pre-processing in a patch-based training increases the rate and amplitude of classes imbalance and negative samples. In fact, PDL reflects a solid application of brain inspired

Table 4
Sample mean ± standard deviation and p-values for DSI, JSI, BF, training and prediction times for the focal, Sequential sU-Net.

	GTV		Bladder		Right Femur		Left Femur		Anorectum		Overall	
Loss Function	GDL	PDL	GDL	PDL	GDL	PDL	GDL	PDL	GDL	PDL	GDL	PDL
DSI	**0.72 ± 0.1**	0.71 ± 0.09	0.81 ± 0.14	**0.82 ± 0.11**	0.86 ± 0.05	**0.87 ± 0.05**	0.87 ± 0.04	**0.88 ± 0.04**	0.58 ± 0.21	**0.6 ± 0.18**	0.77 ± 0.12	**0.78 ± 0.12**
p-value	5.23×10^{-1}		2.05×10^{-1}		5.48×10^{-1}		3.22×10^{-1}		1.67×10^{-1}		N/A	
JSI	**0.57 ± 0.12**	0.56 ± 0.1	0.69 ± 0.16	**0.71 ± 0.14**	0.76 ± 0.07	**0.77 ± 0.08**	0.78 ± 0.07	**0.79 ± 0.06**	0.43 ± 0.19	**0.45 ± 0.17**	0.64 ± 0.14	**0.66 ± 0.15**
p-value	4.95×10^{-1}		2.08×10^{-1}		5.04×10^{-1}		3.24×10^{-1}		2.30×10^{-1}		N/A	
BF	**0.82 ± 0.1**	0.80 ± 0.11	0.9 ± 0.14	**0.92 ± 0.08**	0.92 ± 0.21	**0.96 ± 0.04**	0.94 ± 0.17	**0.97 ± 0.05**	0.64 ± 0.2	**0.71 ± 0.2**	0.86 ± 0.12	**0.87 ± 0.11**
p-value	6.37×10^{-1}		3.46×10^{-1}		4.24×10^{-1}		6.00×10^{-1}		2.21×10^{-1}		N/A	
Training Time in minutes	65.7 ± 0.46	**62.5 ± 0.62**	65.7 ± 0.46	**62.5 ± 0.62**	**133.4 ± 7.85**	135.0 ± 9.78	**133.4 ± 7.85**	135.0 ± 9.78	68.7 ± 3.94	**65.1 ± 0.27**	Mean/ Step[a] 89.26 ± 8.61	**Mean/ Step[a] 87.5 ± 11.5**
p-value	$\mathbf{2.24 \times 10^{-8}}$		$\mathbf{2.24 \times 10^{-8}}$		5.78×10^{-1}		5.78×10^{-1}		$\mathbf{7.02 \times 10^{-3}}$		N/A	
Prediction Time in seconds	3.12 ± 0.69	**2.27 ± 0.34**	3.12 ± 0.69	**2.27 ± 0.34**	**1.97 ± 0.14**	2.01 ± 0.16	**1.97 ± 0.14**	2.01 ± 0.16	3.13 ± 0.27	**2.83 ± 0.79**	Mean/cont[b] 2.18 ± 0.79	**Mean/cont[b] 1.9 ± 0.75**
p-value	$\mathbf{2.40 \times 10^{-3}}$		$\mathbf{2.40 \times 10^{-3}}$		5.08×10^{-1}		5.08×10^{-1}		1.815×10^{-1}		N/A	

The best value achieved (highest for DSI, JSI, and BF) are indicated in bold. P-Values <0.05 are also indicated in bold.
[a] Mean time for 1 sU-Net training.
[b] Mean time for 1 contour segmentation.

Table 5
Total Training and Prediction Times for Classic U-Net and Focal, Sequential sU-Net.

Loss Function	U-Net Classic Approach		sU-Net Focal and Sequential Approach	
	GDL	PDL	GDL	PDL
Total Training Time (min)	**126.30 ±0.49**	129.75 ±4.19	268 ± 8.38	**262 ± 9.92**
Total Prediction Time(sec)	**1.62 ±0.03**	2.04 ±0.06	10.19 ± 0.75	**9.12 ± 0.88**

The best value achieved (lowest for training and prediction times) are indicated in bold.

mathematics, and its potential in semantic segmentation should be explored in different clinical scenarios, anatomical sites, and imaging modalities. While the training time with sU-Net is almost double the training time with classic U-Net, this time is uniquely offline and does not affect the clinical applicability of the method. Similarly, the difference between sU-Net and U-Net prediction time is only of around 10 s, and does not provide any significant clinical advantage of a method to other. It is also worth mentioning that this study was retrospective and data acquisitions were not specifically optimized for the contouring task. Furthermore, our patient sample exhibits a considerable variety of age, histology, and stages. As such, the fact that our method achieved an accurate result in these settings makes it inherently minimally affected by the image artifacts of the general clinical images.

Although five structures were automatically contoured in this experiment, our study has some limitations such as the sub-optimal DSI for the anorectum, the absence of planning target volumes, and the relatively low number of datasets. In future work, tumor volumes should include –in addition to the GTV- the clinical and the planning target volumes. Second, the bowel bag -considered as OAR-was not contoured and providing its automatic contour while improving the anorectum DSI should be addressed. Finally, the number of datasets used in this study is relatively limited and should be expanded and include data from different centers in order to verify robustness before routine clinical deployment and improve the DSI. Depending on the outcomes, sU-Net modifications can be undertaken such as incorporation of attention layers where training [77,78] can be further focalized on some structures of interest.

Regardless of the need for future validation studies with large data, the current approach can be directly implemented in 2 clinical scenarios. In traditional EBRT, our workflow can serve as a quick initial guess that the treating radiation oncologist should review and possible edit before generating the dosimetric plan. In this case, our prediction and preprocessing times are in the order of seconds, and at least half of the automatic contours would not need any significant adjustment. Even if the automatically generated contours need manual tuning, this is expected to be quicker than de-novo contouring. As such, we speculate that at least 50% of the contouring time can be reduced. The second clinical scenario is MR-based online adaptive radiation therapy. In this setting an MR is acquired immediately before radiation treatment delivery while the patient is on the treatment table. Afterwards, one or more physicians will try to rapidly perform contouring while the patient remains on the treatment table [25]. Manual contouring in MR-based online adaptive radiation therapy consumes simultaneously the patient's time, in addition to a significant human (physician, radiation physicist, radiation technologist) and technological resources (adaptive radiation therapy platform). In such a setting, computer aided contouring would dramatically accelerate the process by cost-effectively reducing the time needed for contouring as -similarly to EBRT-our contours can serve as a quick initial guess.

5. Conclusion

A novel deep-learning approach for multi-class automatic contouring in FDG-PET/MR based cervical cancer has been investigated. The approach imitates the model of human thinking by leveraging anatomic and imaging properties to the efficiency of semantic segmentation with a new, shallow U-net architecture. sU-Net has a smaller number of parameters to be optimized compared to U-Net, thus more suitable for tasks where a limited set of data is available. However, the results with the shallow structure are satisfactory as the overall DSI was is 0.7. The suggested workflow presents the major advantage of much shorter contouring time compared with manual delineation, and should be validated in clinical settings with further data and dosimetric studies.

Funding

This research was supported in part by National Cancer Institute of the National Institute of Health, USA, under award number R01CA196687. The authors are thankful for Miss Taniya Parker (Department of Radiology, School of Medicine, Case Western Reserve University, Cleveland, OH through the YES award R25CA221718) for her technical assistance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ibmed.2021.100026.

References

[1] Arbyn M, Weiderpass E, Bruni L, de Sanjosé S, Saraiya M, Ferlay J, et al. Estimates of incidence and mortality of cervical cancer in 2018: a worldwide analysis. The Lancet Global Health 2020;8:e191–203.

[2] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA A Cancer J Clin 2020;70:7–30.

[3] Health NCIatNIo. Cervical cancer treatment (PDQ®)–Health professional version. 2016.

[4] Xu Z, Traughber BJ, Fredman E, Albani D, Ellis RJ, Podder TK. Appropriate methodology for EBRT and HDR intracavitary/interstitial brachytherapy dose composite and clinical plan evaluation for patients with cervical cancer. Practical radiation oncology 2019;9:e559–71.

[5] Song Y, Erickson B, Chen X, Li G, Wu G, Paulson E, et al. Appropriate magnetic resonance imaging techniques for gross tumor volume delineation in external beam radiation therapy of locally advanced cervical cancer. Oncotarget 2018;9:10100.

[6] Grosu A-L, Nieder C. Target volume definition in radiation oncology. Springer; 2015.

[7] Koh W-J, Abu-Rustum NR, Bean S, Bradley K, Campos SM, Cho KR, et al. Cervical cancer, version 3.2019, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 2019;17:64–84.

[8] Balleyguier C, Sala E, Da Cunha T, Bergman A, Brkljacic B, Danza F, et al. Staging of uterine cervical cancer with MRI: guidelines of the European Society of Urogenital Radiology. Eur Radiol 2011;21:1102–10.

[9] Rauch GM, Kaur H, Choi H, Ernst RD, Klopp AH, Boonsirikamchai P, et al. Optimization of MR imaging for pretreatment evaluation of patients with endometrial and cervical cancer. Radiographics 2014;34:1082–98.

[10] Patel MR, Klufas RA, Alberico RA, Edelman RR. Half-fourier acquisition single-shot turbo spin-echo (HASTE) MR: comparison with fast spin-echo MR in diseases of the brain. Am J Neuroradiol 1997;18:1635–40.

[11] Hennig J, Nauerth A, Friedburg H. RARE imaging: a fast imaging method for clinical MR. Magn Reson Med 1986;3:823–33.

[12] Ma X, Tian J, Jiang G, Liang L, Zeng S, Li W. A pilot study on the application of FFE and SSh-TSE sequences in ocular MRI. Eye Sci 2011;26:173–9.

[13] Terezakis SA, Heron DE, Lavigne RF, Diehn M, Loo Jr BW. What the diagnostic radiologist needs to know about radiation oncology. Radiology 2011;261:30–44.

[14] Gay HA, Barthold HJ, O'Meara E, Bosch WR, El Naqa I, Al-Lozi R, et al. Pelvic normal tissue contouring guidelines for radiation therapy: a Radiation Therapy Oncology Group consensus panel atlas. Int J Radiat Oncol Biol Phys 2012;83:e353–62.

[15] Rash DL, Lee YC, Kashefi A, Durbin-Johnson B, Mathai M, Valicenti R, et al. Clinical response of pelvic and para-aortic lymphadenopathy to a radiation boost in the definitive management of locally advanced cervical cancer. Int J Radiat Oncol Biol Phys 2013;87:317–22.

[16] Zhang S, Xin J, Guo Q, Ma J, Ma Q, Sun H, et al. Defining PET tumor volume in cervical cancer with hybrid PET/MRI: a comparative study. Nucl Med Commun 2014;35:712–9.

[17] Zhang S, Xin J, Guo Q, Ma J, Ma Q, Sun H. Comparison of tumor volume between PET and MRI in cervical cancer with hybrid PET/MR. Int J Gynecol Canc 2014;24:744–50.

[18] Zincirkeser S, Şahin E, Halac M, Sager S. Standardized uptake values of normal organs on 18F-fluorodeoxyglucose positron emission tomography and computed tomography imaging. J Int Med Res 2007;35:231–6.

[19] Aselmaa A, van Herk M, Song Y, Goossens RH, Laprie A. The influence of automation on tumor contouring. Cognit Technol Work 2017;19:795–808.

[20] Vorwerk H, Zink K, Schiller R, Budach V, Böhmer D, Kampfer S, et al. Protection of quality and innovation in radiation oncology: the prospective multicenter trial the German Society of Radiation Oncology (DEGRO-QUIRO study). Strahlenther Onkol 2014;190:433–43.

[21] Scardapane A, Lorusso F, Scioscia M, Ferrante A, Ianora AAS, Angelelli G. Standard high-resolution pelvic MRI vs. low-resolution pelvic MRI in the evaluation of deep infiltrating endometriosis. Eur Radiol 2014;24:2590–6.

[22] Kim N, Chang JS, Kim YB, Kim JS. Atlas-based auto-segmentation for postoperative radiotherapy planning in endometrial and cervical cancers. Radiat Oncol 2020;15:1–9.

[23] Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A review of methods of analysis in contouring studies for radiation oncology. Journal of medical imaging and radiation oncology 2010;54:401–10.

[24] Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. Physics and Imaging in Radiation Oncology 2020;13:1–6.

[25] Liang F, Qian P, Su K-H, Baydoun A, Leisser A, Van Hedent S, et al. Abdominal, multi-organ, auto-contouring method for online adaptive magnetic resonance guided radiotherapy: an intelligent, multi-level fusion approach. Artif Intell Med 2018;90:34–41.

[26] Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, 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 2018;50:13–9.

[27] Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. Radiother Oncol 2018;126:312–7.

[28] Cardenas CE, McCarroll RE, Court LE, Elgohari BA, Elhalawani H, Fuller CD, 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 2018;101:468–78.

[29] Capelle A-S, Alata O, Fernandez C, Lefèvre S, Ferrie J. Unsupervised segmentation for automatic detection of brain tumors in MRI. In: Proceedings 2000 international conference on image processing (Cat No 00CH37101). IEEE; 2000. p. 613–6.

[30] Vinitski S, Gonzalez CF, Knobler R, Andrews D, Iwanaga T, Curtis M. Fast tissue segmentation based on a 4D feature map in characterization of intracranial lesions. J Magn Reson Imag: An Official Journal of the International Society for Magnetic Resonance in Medicine 1999;9:768–76.

[31] Zhang Y, Li T, Xiao H, Ji W, Guo M, Zeng Z, et al. A knowledge-based approach to automated planning for hepatocellular carcinoma. J Appl Clin Med Phys 2018;19:50–9.

[32] Hwee J, Louie AV, Gaede S, Bauman G, D'Souza D, Sexton T, et al. Technology assessment of automated atlas based segmentation in prostate bed contouring. Radiat Oncol 2011;6:110.

[33] Ghosh S, Das N, Das I, Maulik U. Understanding deep learning techniques for image segmentation. ACM Comput Surv 2019;52:1–35.

[34] Boldrini L, Bibault J-E, Masciocchi C, Shen Y, Bittner M-I. Deep learning: a review for the radiation oncologist. Frontiers in Oncology 2019;9:977.

[35] Fukushima K. Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol Cybern 1980;36:193–202.

[36] Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. 2020. arXiv preprint arXiv:200105566.

[37] Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: 2017 international conference on engineering and technology (ICET). IEEE; 2017. p. 1–6.

[38] Qian P, Xu K, Wang T, Zheng Q, Yang H, Baydoun A, et al. Estimating CT from MR abdominal images using novel generative adversarial networks. J Grid Comput 2020:1–16.

[39] Philbrick KA, Weston AD, Akkus Z, Kline TL, Korfiatis P, Sakinis T, et al. RIL-contour: a medical imaging dataset annotation tool for and with deep learning. J Digit Imag 2019;32:571–81.

[40] Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. Radiographics 2017;37:2113–31.

[41] Ronneberger O, Fischer P, Brox T. Dental X-ray image segmentation using a U-shaped Deep Convolutional network. 2015. ISBI, http://www.ontustedutw/~cweiwang/ISBI2015/challenge2/isbi2015_Ronnebergerpdf. [Accessed 16 May 2019].

[42] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. Springer; 2016. p. 424–32.

[43] Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imag 2014;34:1993–2024.

[44] Dong H, Yang G, Liu F, Mo Y, Guo Y. Automatic brain tumor detection and segmentation using u-net based fully convolutional networks. In: Annual conference on medical image understanding and analysis. Springer; 2017. p. 506–17.

[45] Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys 2019;46:2157–68.

[46] Balagopal A, Kazemifar S, Nguyen D, Lin M-H, Hannan R, Owrangi A, et al. Fully automated organ segmentation in male pelvic CT images. Phys Med Biol 2018;63:245015.

[47] Chino J, Annunziata CM, Beriwal S, Bradfield L, Erickson BA, Fields EC, et al. Radiation therapy for cervical cancer: executive summary of an ASTRO clinical practice guideline. Practical Radiation Oncology 2020;10(4):220–34.

[48] Chen L, Shen C, Zhou Z, Maquilan G, Albuquerque K, Folkert MR, et al. Automatic PET cervical tumor segmentation by combining deep learning and anatomic prior. Phys Med Biol 2019;64:085019.

[49] Lin Y-C, Lin C-H, Lu H-Y, Chiang H-J, Wang H-K, Huang Y-T, et al. Deep learning for fully automated tumor segmentation and extraction of magnetic resonance radiomics features in cervical cancer. Eur Radiol 2020;30:1297–305.

[50] Liu Z, Liu X, Xiao B, Wang S, Miao Z, Sun Y, et al. Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network. Phys Med 2020;69:184–91.

[51] Luc P, Couprie C, Chintala S, Verbeek J. Semantic segmentation using adversarial networks. 2016. arXiv preprint arXiv:161108408.

[52] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.

[53] Rezaei M, Yang H, Meinel C. Conditional generative refinement adversarial networks for unbalanced medical image semantic segmentation. 2018. arXiv preprint arXiv:181003871.

[54] Sudre CH, Li W, Vercauteren T, Ourselin S, Cardoso MJ. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Deep learning in medical image analysis and multimodal learning for clinical decision support. Springer; 2017. p. 240–8.
[55] O'Shea K, Nash R. An introduction to convolutional neural networks. 2015. arXiv preprint arXiv:151108458.
[56] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. 2015. arXiv preprint arXiv:150203167.
[57] Hahnloser RH, Seung HS. Permitted and forbidden sets in symmetric threshold-linear networks. Adv Neural Inf Process Syst 2001:217–23.
[58] Christlein V, Spranger L, Seuret M, Nicolaou A, Král P, Maier A. Deep generalized max pooling. 2019. arXiv preprint arXiv:190805040.
[59] Andrearczyk V, Whelan PF. Convolutional neural network on three orthogonal planes for dynamic texture classification. Pattern Recogn 2018;76:36–49.
[60] Sudholt S, Fink GA. PHOCNet: a deep convolutional neural network for word spotting in handwritten documents. In: 2016 15th international conference on frontiers in handwriting recognition (ICFHR). IEEE; 2016. p. 277–82.
[61] Nagi J, Ducatelle F, Di Caro GA, Cireşan D, Meier U, Giusti A, et al. Max-pooling convolutional neural networks for vision-based hand gesture recognition. In: 2011 IEEE international conference on signal and image processing applications (ICSIPA). IEEE; 2011. p. 342–7.
[62] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer; 2015. p. 234–41.
[63] Li Y, Zhang T, Liu Z, Hu H. A concatenating framework of shortcut convolutional neural networks. 2017. arXiv preprint arXiv:171000974.
[64] Qian P, Chen Y, Kuo J-W, Zhang Y-D, Jiang Y, Zhao K, et al. mDixon-based synthetic CT generation for PET attenuation correction on abdomen and pelvis jointly using transfer fuzzy clustering and active learning-based classification. IEEE Trans Med Imag 2019;39(4):819–32.
[65] Wang L, Wang C, Sun Z, Chen S. An improved dice loss for pneumothorax segmentation by mining the information of negative areas. IEEE Access 2020;8:167939–49.
[66] Zadeh LA. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Set Syst 1997;90:111–27.
[67] Zaidi H, Ojha N, Morich M, Griesmer J, Hu Z, Maniawski P, et al. Design and performance evaluation of a whole-body Ingenuity TF PET–MRI system. Phys Med Biol 2011;56:3091.
[68] Kalemis A, Delattre BM, Heinzer S. Sequential whole-body PET/MR scanner: concept, clinical use, and optimisation after two years in the clinic. The manufacturer's perspective. Magnetic Resonance Materials in Physics, Biology and Medicine 2013;26:5–23.
[69] Muzic RF, Cornelius S. COMKAT: compartment model kinetic analysis tool. J Nucl Med 2001;42:636–45.
[70] Fang Y-HD, Asthana P, Salinas C, Huang H-M, Muzic RF. Integrated software environment based on COMKAT for analyzing tracer pharmacokinetics with molecular imaging. J Nucl Med 2010;51:77–84.
[71] Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. arXiv preprint arXiv:14126980.
[72] Jaccard P. Étude comparative de la distribution florale dans une portion des Alpes et des Jura. Bull Soc Vaudoise Sci Nat 1901;37:547–79.
[73] Csurka G, Larlus D, Perronnin F, Meylan F. What is a good evaluation measure for semantic segmentation? BMVC2013. 2013.
[74] Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. 2011.
[75] Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. Acad Radiol 2004;11:178–89.
[76] Wang J, Lu J, Qing G, Shen L, Sun Y, Ying H, et al. A novel deep learning based auto segmentation for rectum tumor on MRI image. Int J Radiat Oncol Biol Phys 2018;102:e548.
[77] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: learning where to look for the pancreas. 2018. arXiv preprint arXiv:180403999.
[78] Sinha A, Dolz J. Multi-scale self-guided attention for medical image segmentation. IEEE Journal of Biomedical and Health Informatics 2021 Jan;25(1).