Segmentation of deformed kidneys and nephroblastoma using Case-Based Reasoning and Convolutional Neural Network

Florent Marie, Lisa Corbat, Yann Chaussy, Thibault Delavelle, Julien Henriet, Jean-Christophe Lapayre

Segmentation of deformed kidneys and nephroblastoma using Case-Based Reasoning and Convolutional Neural Network

Florent Marie[a,*], Lisa Corbat[a], Yann Chaussy[b], Thibault Delavelle[a], Julien Henriet[a], Jean-Christophe Lapayre[a]

[a] *FEMTO-ST Institute, DISC, CNRS, Univ. Bourgogne-Franche-Comté, 16 route de Gray, 25030 Besançon, France*
[b] *CHRU Jean Minjoz, 3 boulevard Fleming, 25030 Besançon, France*

Abstract

Most often, image segmentation is not fully automated and a user is required to lead the process in order to obtain correct results. In a medical context, segmentation can furnish a lot of information to surgeons, but this task is rarely executed. Artificial Intelligence (AI) is a powerful approach for devising a viable solution to fully automated treatment. In this paper, we have focused on kidneys deformed by nephroblastoma. Yet, a frequent medical constraint is encountered which is a lack of data with which to train our system. In function of this constraint, two AI approaches were used to segment these structures. First, a Case Based Reasoning (CBR) approach was defined which can enhance the growth of regions for segmentation of deformed kidneys with an adaptation phase to modify coordinates of recovered seeds. This CBR approach was confronted with manual region growing and a Convolutional Neural Network (CNN). The CBR system succeeded in performing the best segmentation for the kidney with a mean Dice of 0.83. Deep Learning was then examined as a possible solution, using the latest performing networks for image segmenta-

*Corresponding author, email adress: florent.marie@univ-fcomte.fr
Email addresses: florent.marie@univ-fcomte.fr (Florent Marie), lisa.corbat@univ-fcomte.fr (Lisa Corbat), yann.chaussy@univ-fcomte.fr (Yann Chaussy), thibault.delavelle@edu.univ-fcomte.fr (Thibault Delavelle), julien.henriet@univ-fcomte.fr (Julien Henriet), jean-christophe.lapayre@univ-fcomte.fr (Jean-Christophe Lapayre)

tion. However, for relevant efficiency, this method requires a large data set. An option would be to manually segment only certain representative slices from a patient and, using them, to train a Convolutional Neural Network how to segment. In this article the authors propose an evaluation of a CNN for medical image segmentation following different training sets with avariable number of manual segmentations. To choose slices to train the CNN, an Overlearning Vector for Valid Sparsed SegmentatIONs (OV^{2}ASSION) was used, with the notion of gap between two slices from the training set. This protocol made it possible to obtain reliable segmentations of tumor per patient with a low data set and to determine that only 26% of initial segmented slices are required to obtain a complete segmentation of a patient with a mean Dice of 0.897.

Keywords: Case-Based Reasoning, Convolutional Neural Network, Segmentation, Tumor, Healthcare imaging, Artificial Intelligence

1. Introduction

Nephroblastoma, also called Wilms tumor, is one of the most frequent abdominal tumors observed in children (generally in both boys and girls, 1 to 5 years of age), representing 5 to 14% of malignant pedriatric tumors. As indicated by its name, this type of tumor is situated in the kidney. Most often, its initial diagnosis is based on imaging. Generally, ultrasound observations are first planned in order to confirm the tumor's existence and to approximate its position. A medical scan then locates it with greater accuracy, along with affected organs and healthy tissues. Radiologists and surgeons need 3-Dimensional (3D) representations of the tumor and the border organs in order to establish the diagnosis, to plan the surgery (estimated quantity of blood, specialized equipment needed, estimation of the duration of the surgery, etc.) and also in order to lead the surgeon's actions during surgery.

Segmentation is one of the key steps in the construction of such a 3D representation. During this process, each pixel of all scans has to be affected to one and only one region. Each region represents a given structure (right or left

kidney, medulas, tumors, muscles, veins, cavities, etc.). The problem resides in the unforeseeable nature of the situation of the kidneys, their shapes and areas since the area covered by the tumor is totally unpredictable and may differ from one patient to another. Consequently, radiologists and surgeons must lead and verify the segmentations of more than 200 scans manually, which, in practice, is out of the question. Artificial Intelligence (AI) is a powerful tool that may provide a viable solution for a fully automated treatment.

Our ambition is double. First, we want to introduce a process based on expert knowledge and experience in order to lead the segmentation process. We have thus considered the Case-Based Reasoning (CBR) approach to be suited to this particular type of problem. This paper presents a CBR system which can enhance segmentation of kidneys deformed by nephroblastoma. Second, Deep Learning would be also an interesting solution with the latest performing networks for image segmentation, but only if the difficulties of data lacking are overcome. We also discuss a method used to train CNNs from a small training set (samples) in order to segment the totality of a patient's abdomen.

The originality of our CBR approach resides in the way our CBR system adapts solutions from past situations. The principle of the region-growing segmentation consists in placing seeds at different points of the image to be segmented, and then to verify whether the pixels around these seeds have grey levels close enough to be integrated into the same region following a local threshold (difference between the candidate pixel and its neighborhood) and a global threshold (difference between the candidate pixel and the regional mean). This algorithm allows each region to grow until each pixel of the picture in question belongs to one and only one region. The main problem with the region-growing method is that a user needs to manually place the seeds in the image. Automated methods exist but are not efficient enought every time and often produce an over-segmentation. Our idea consists of using a CBR system which places the seeds in the right places, and to then perform a region-growing segmentation. In addition, our system can modify the position of a seed during the adaptation phase in order to choose a pixel which is a better match for the expected grey

level of a kidney.

The main limitation of Deep Learning is the size of data for the training in order to keep a good generalization capacity. In a medical context, this limitation represents one of the main obstacles to overcome. Rather than trying to train the network on a representative training set, the idea is to perform an overlearning with somes slices, manually segmented by a surgeon, from a patient we want to segment. Yet, we need to determine the number and position of required manual segmentations for relevant results. We introduce notions of vector and gap to select slices, perform the training and evaluate computed outputs thanks to a protocol called Overlearning Vector for Valid Sparsed SegmentatIONs (OV^2 ASSION).

Following this introduction, a brief related work section is realized about images segmentation with AI approaches. In the third part, the CBR system we have designed is presented in order to compensate for the unpredictability of kidney tumor shapes: its case representation and all the phases of the system, accentuating the adaptation phase. This paper also presents our gap CNN training method. In the fourth part, the performances and evaluations of our methods are presented (CBR and gap approach); these results are then discussed regarding our general purpose.

2. Related work

Many research studies relative to segmentation enhanced by AI are found in the literature. Indeed, current segmentation methods are enhanced by CBR (Perner (1999, 2001); Frucci et al. (2008)), genetic algorithms (Golobardes et al. (2002)), knowledge stored in ontologies (Colliot et al. (2006); Hudelot et al. (2008); Trzupek et al. (2011); Burgos-Artizzu et al. (2009)), Markov random fields (Kato et al. (2012)) and with Deep learning (Litjens et al. (2017)).

Deep learning (like Convolutional Neural Network (CNN)) is one of the most widely experimented, efficient and promising tools. The term is actually a generic name which refers to a set of tools that come from Artificial Neural

Networks (ANN). The first Fully Convolutional Network (FCN) was designed in 2015 by Long et al. (2015) and gives impressive performances. E-Net (Efficient Neural Network) in Paszke et al. (2016) and SegNet in Badrinarayanan et al. (2015) are convolutional networks developed in order to perform segmentation of indoor scenes and road scenes in real-time. DeconvNet (Deconvolution Network) is a CNN whose principle consists of aggregating a convolutional network to a deconvolutional network (Noh et al. (2015)), while DecoupledNet (Hong et al. (2015)) is similar but with bridging layers serving as intermediaries between the two networks. Thong et al. (2016) used CNN in order to perform segmentation of healthy kidneys. Their network is not fully convolutional and classifies each pixel separately via a window. They moved this window along the image to segment all parts of the kidney. Zhou et al. (2017) suggest a 3D segmentation of CT-scans from 2D slices in three directions by means of a fuzzy localization of organs (bounding box), a segmentation of each 2D slice and a majority voting. The results were good, but no attempt was made to segment a tumoral kidney, thus presenting a difficulty. Actually, convolutional networks must first be trained to recognize the shapes of these organs. Since this study was focused on healthy organs, their shapes and areas are more or less the same from one subject to another. In contrast, our study aims at performing segmentation of tumoral kidneys with unpredictable shapes and situations, with very different forms from one patient to another. In the biology field, the most widely used FCN for these purposes is U-Net by Ronneberger et al. (2015). A 3D variation of U-Net was proposed in 2016 Çiçek et al. (2016) and evaluated on Xenopus kidney embryos from confocal microscopy images. The Originality of this paper resides in the training method used, with a few slices annotated by the user, prior to a complete segmentation calculation with a CNN during the second phase.

Brain tumor images from the 2015 Brain Tumor Segmentation Challenge (BRATS) were segmented by DeepMedic network, obtaining a mean Dice of 0.898 with Conditional Random Field (CRF) post-processing Kamnitsas et al. (2017). More recently, Xia & Kulis (2017) designed W-Net, an unsupervised CNN based on U-Net, inside an encoder-decoder network, with CRF post-

processing. An unsupervised method would be a great advantage for medical applications, where image labeling is a critical issue. Results are positive but the images used were much simpler than medical images. Panigrahi et al. (2018) used in 2018 a fuzzy c-means clustering combined with multi-scale vector field convolution in order to segment breast tumors in ultrasound images. Lim & Mandava (2018) suggested a semi-automatic approach for segmentation of BRATS images but only with an average Dice accuracy of 0.7. Considering MRI brain images have more contrasts and less noise than CT-Scans, this approach does not seem relevant in our context.

Though, in recent studies, Deep learning appears to give the most accurate results, this technique requires a lot of data in order to be trained. We can also find some studies with an unsupervied learning. In contrast, CBR gives an advantage to knowledge and enriches itself following its experiments (Kolodner (2014)). A large number of CBR systems designed for Health Science (CBR-HS) can also be found in Diaz et al. (2006); Montani (2009); Marling et al. (2014); Perner & Attig (2011); Attig & Perner (2012); Henriet et al. (2014b); Henriet & Lang (2014). In particular, Perner (1999) designed a system for segmentation of brain images with a cut histogram method. Frucci et al. (2008) adapted and improved this system with a watershed method. We chose to implement a CBR system with a region-growing method but, in order to place seeds in the optimum position, we needed an adaptation phase.

Many adaptation strategies can be found in the literature. Adaptation by generalization/specialization requires a hierarchical organization of the CBR source cases according to generalization/specialization relations. Some characteristics are hidden in the generalization process whereas special ones are added to the general case during the specialization process. Adaptation using Adaptation Rules (Melis et al. (1998)) consists of computing a solution for a target case by applying a function which takes as parameters the target case, a source case that presents some similarities and its solution. Differential Adaptation (Fuchs et al. (2000)) is based on the evaluation of the variations between the source and target cases: an approximate solution to the target case is computed by

applying the variations between the two cases to the solution for the source case under consideration. Conservative Adaptation (Lieber (2007)) is based on the Revision Theory which considers knowledge updates. This kind of adaptation consists of minimizing the modifications to be applied to the knowledge and has been applied to the resolution of spatial and temporal problems in Dufour-Lussier et al. (2012) and also to oncology in d'Aquin et al. (2006). Cordier et al. (2006) used Influence functions that link variations in problem descriptors to those in solution descriptors. In the CBR-HS EquiVox, an adaptation based on rules defined by experts' experiences and Artificial Neural Networks (ANN) has been implemented and enhanced by a precision combination vector (Henriet & Chatonnay (2013); Henriet et al. (2014a)). In the present study, the neighborhoods of the seeds are explored in order to match the desired grey levels as much as possible. Consequently, our adaptation is a kind of Conservative Adaptation led by rules.

3. Materials and Methods

Two methods are presented in this section. First, the CBR system associated with a region growing process is explained. The CNN training method from a small data set for tumor segmentation (OVÂ²ASSION) is then presented.

3.1. CBR system

This part of the paper presents the CBR system defined for image segmentation with the region-growing method, as summarized in Figure 1. All the CBR phases are explained below. Our system comes from a CT-scan and seeks the closest stored image already segmented in the case base. It calculates a similarity value for each stored case and, during a retrieval phase, extracts the source case with the highest similarity. Then, extracted parameters of segmentation are adapted to the current case via an adaptation phase. These adapted parameters are used to perform a new segmentation. Finally, the result is evaluated by an expert and stored in the case base as a new source case if the segmentation is relevant.

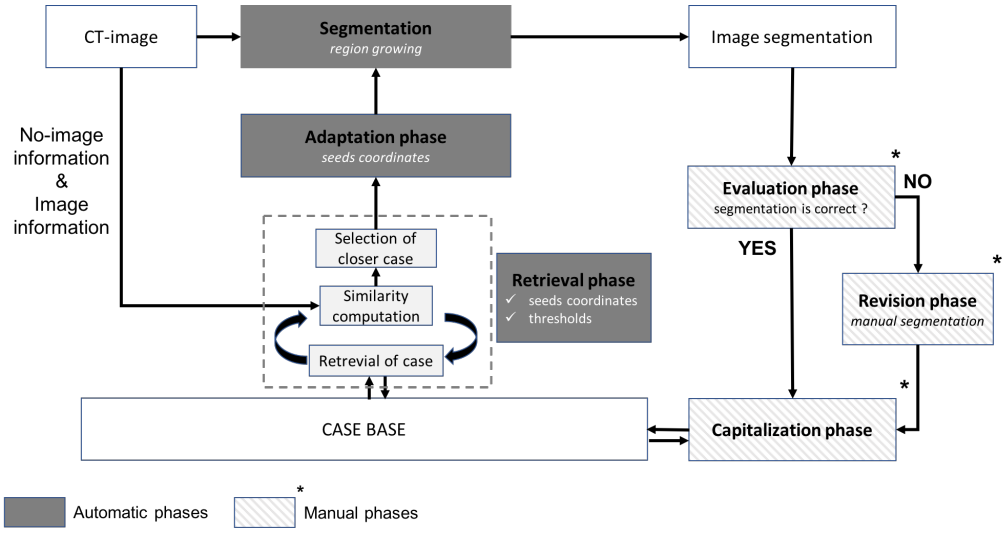

Figure 1: Overview of our CBR system

3.1.1. Case model approach

A case is composed of two parts: the *Problem* and the *Solution*. The *Problem* part describes the characteristics of the problem to be solved, and the *Solution* part provides the way to solve it. The case model of CBR is described in Figure 2, with the left side representing the *Problem* and the right side representing the *Solution*.

In our study, the *Problem* part has to describe the CT-images, i.e. descriptors that provide information about the structure or the statistics of the image. The *Solution* part should give the seed locations and the thresholds for the region-growing algorithm.

As in Attig and Perner's approach (Perner & Attig (2011); Attig & Perner (2012, 2009)), the *Problem* part of our cases is composed of:

- meta-data representing the information of the patient: sex, age, height and weight;

- statistical image information: mean, kurtosis, skewness and variance;

The *Solution* part of our cases contains the values of the thresholds (local and global) for each structure to segment, a list of positions for the seeds and a list of pretreatments (giving the order of the filters to apply and their parameters). We use three different pretreatments in order to enhance the contrast of the images: histogram equalization, median filter and unsharp mask. Each image has its own characteristics, thus it is necessary to add the preprocessing to the *Solution part*, in order to fit each case.

$$Case = \begin{bmatrix} patient\ sex \\ patient\ age \\ patient\ height \\ patient\ weight \\ image\ mean \\ image\ kurtosis \\ image\ skewness \\ image\ variance \end{bmatrix} + \begin{bmatrix} thresholds\ for\ each\ structure \\ list\ of\ pretreatments \\ 2D\ coordinates\ of\ kidney\ seeds \end{bmatrix}$$

Description of problem part **Description of solution part**

Figure 2: The CBR case model: problem part and solution part

3.1.2. Retrieval phase

This section describes how the retrieval phase is performed using a similarity calculation between the stored cases.

As explained in the previous section, our case base is composed of two types of data: meta-data and image characteristics (see section 3.1.1). We tested two variants of similarity formula between two images x and y. First, we used Perner's formula as in Perner (1999), called $s_1(x, y)$ with a component $SM_d(x, y)$ for meta-data (a derivation of Tversky index) and another $SI(x, y)$ for image information. The second is a hybrid formula called $s_2(x, y)$. It takes back the meta-data formula of Perner $SM_d(x, y)$ but uses MSSIM (Mean Struc-

tural SIMilarity) from Wang et al. (2004) criteria ($MSSIM(x, y)$) for the image characteristics. The MSSIM criteria is an improvement of the SSIM (Structural Similarity) criteria that is commonly used in image compression. MSSIM uses an iterative windowing to increase the capacity of structural comparison on images (i.e. each window describes an ROI (Region Of Interest) which is compared independently). The following formulas show the construction of both of the similarity calculations between two images x and y ($s_1(x, y)$ and $s_2(x, y)$):

$$s_1(x, y) = \frac{1}{2}(SM_d(x, y) + SI(x, y)) \tag{1}$$

$$s_2(x, y) = \frac{1}{2}(SM_d(x, y) + MSSIM(x, y)) \tag{2}$$

The meta-data component is computed as follows:

$$SM_d(x, y) = \frac{|A_i|}{\alpha|A_i| + \beta|D_i| + \gamma|E_i|} \tag{3}$$

where A_i is the number of common features between x and y, D_i the features only in x and E_i the features only in y.α, β and γ are weight factors such as $\alpha = \beta = 1$ and $\gamma = 0.5$. Weight values come from Perner (1999).

The image component in $s_1(x, y)$ is:

$$SI(x, y) = \frac{1}{K}\sum_{i=1}^{K} w_i|\frac{C_{ix} - C_{imin}}{C_{imax} - C_{imin}} - \frac{C_{iy} - C_{imin}}{C_{imax} - C_{imin}}| \tag{4}$$

C_{ix} and C_{iy} are the ith feature of both of the images x and y. C_{imin} and C_{imax} are the minimum and the maximum of the ith feature in the case base, respectively. w_i is a weight factor which allows to ponder each image features. In our study, all weight factors are set to 1.

The image component in $s_2(x, y)$ is:

$$MSSIM(x, y) = \frac{1}{M}\sum_{i=1}^{M} SSIM(x_i, y_i) \tag{5}$$

where M is the number of windows and SSIM is

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{6}$$

μ is the mean, σ is the standard deviation and $C_1 = 0.01 * L$ and $C_2 = 0.03 * L$ with $L = 255$.

3.1.3. Adaptation phase

In our system, adaptation consists of correctly positioning the seeds. Since the tumor form and position are unpredictable, kidney position is vulnerable to errors. Moreover, the kidney may be a small region, especially when the tumor crushes it. Consequently, the retrieved case is not always exactly the same as the new case we want to solve. So, seed position has to be adapted, especially for small regions..

We assumed that after the retrieval phase, the position of a seed is inside or close to its dedicated region. After applying the pretreatments, we figured out that the grey-level intensity of an object to segment is almost the same. Therefore, we can automatically infer the correct position of seeds considering the grey-level intensity of the pixel. We defined a coherence interval for each small object to segment, corresponding to an interval of grey-level intensity in which a seed must be situated. With the following test we can define a procedure to verify whether or not a seed belongs to its dedicated region:

$$\begin{gathered} \forall seed,\ isCorrectlyPlaced(seed) = true \\ if\ I(seed) \in CI_{kidney} / CI_{kidney} = [220, 255] \end{gathered} \tag{7}$$

From this test, we defined an algorithm that can place the seed in the correct position on the image (i.e. in a pixel belonging to the kidney). Figure 3 illustrates the iterative extending of the neighbors from the initial seed position. The main idea is to evaluate the seed's neighbors (8-connected), and check their coherence using the test presented above. If no candidate satisfies the test, we iteratively expand the neighbors' scope until we find a coherent pixel. In Figure

3, each square represents a pixel of the image: in the first iteration we evaluate the white seed's direct neighbors, in the second one we increase the scope to the pixels at a distance of two from the seed, and so on.

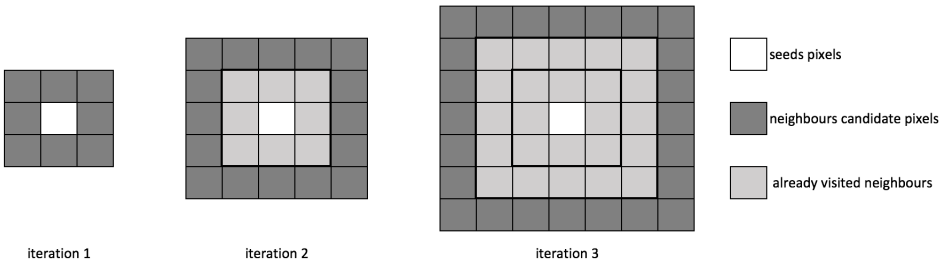

Figure 3: Overview of the evaluated neighbours during the execution of the adaptation algorithm (for the position of the seeds) on 3 iterations

Algorithm 1 presented below gives the details of this process. Giving a seed S and a coherence interval CI, iit returns the 2D coordinate representing the seed's adapted position. For each evaluated neighbor, the test presented above is performed to check whether or not the pixel is coherent, and, at each iteration, the scope (α) is increased.

Figure 4 shows an example of execution of the algorithm. From the initial seed (in white), we expand twice until we reach a coherent candidate (the black pixel), which is part of the kidney region according to its grey-level intensity (> 220).

3.1.4. Revision and Capitalization phases

As presented in Figure 1, all these phases are conducted by an expert. The expert evaluates the segmentation results, visually, and decides whether or not to do it again with a region growing process manually guided. The capitalisation phase is systematic, we enrich the case base after every segmentation obtained. Ideally, a satisfaction measure would determine if a resolved case could be added to the base.

Algorithm 1: Adaptation algorithm for seed position

inputs : a seed S (with its 2D coordinates in the image $S.x$ et $S.y$) and the intensity coherence interval for dedicated region of S : CI

output: a 2D Point describing the new position of the seed S

$\alpha \leftarrow 1$;

while $intensity(S.x, S.y) \notin CI$ **do**

  for i *from* $-\alpha$ *to* α *step of* 2α **do**

    for j *from* $-\alpha$ *to* α *step of* 2α **do**

      if $intensity(S.x + i, S.y + j) \in CI$ **then**

        return the point of coordinates $(S.x + i, S.y + j)$;

      end

    end

  end

  $\alpha + +$;

end

return error "no candidate satisfies the coherence interval";

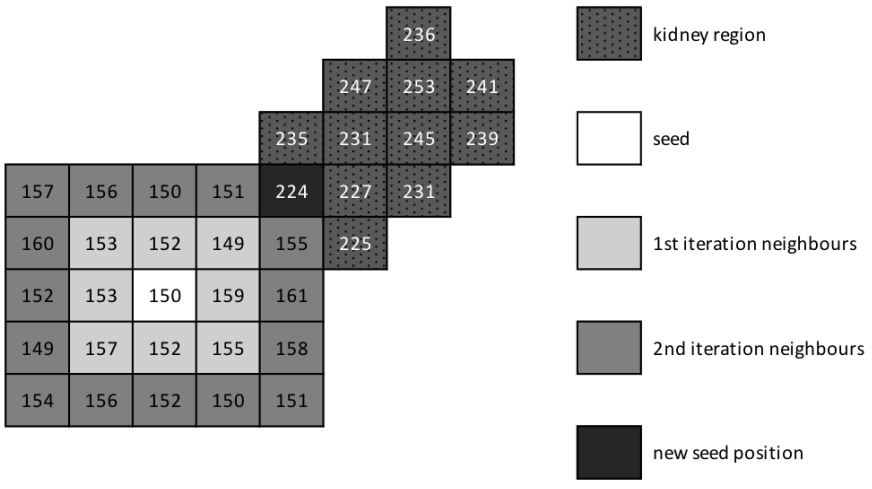

Figure 4: Execution example of the adaptation algorithm to find a new correct seed

3.2. Overlearning Vector for Valid Sparsed SegmentatIONs (OVÅSSION) method for CNN

The problem with CNN resides in the necessity to train them over large data sets. Nevertheless, this is not possible in our case since each tumor is unique and we have a limited number of tumor segmentations from different patients. However in a given patient, the tumor has the same texture and is generally homogenous. We thus explore the possibility of establishing a deep neural network dedicated to and trained for each patient that will be able to carry out the automatic tumor segmentations without human intervention.

As we perform a single segmentation per patient, some initial segmentations of this patient must be integrated into the learning set. The aim is to minimize the number of data in the learning set while optimizing the results of the tumor segmentation so that a maximum of segmentations are calculated from a minimum of initial segmentations. Those initial segmentations may be obtained manually or may be the most accurate segmentation performed using classical segmentation methods. So, this method consists of an overlearning on initial slices before a complete segmentation with a CNN.

We observed only a slight difference in the tumor segmentation from the same patient from one neighboring slice to another. As shown in Figure 5, the quantity of data can be minimized by taking into account only certain manual segmentations at different levels and which will be representative of the neighboring segmentations.

The problem is to determine how many slices, between two chosen slices, can be ignored during training while maintaining correct learning. In other words, what is the optimal gap between two chosen slices ? As shown in Figure 5, we propose to optimize tumor segmentation results for a given patient by minimizing the number of data in the learning set, while maximizing the gap between each chosen slice.

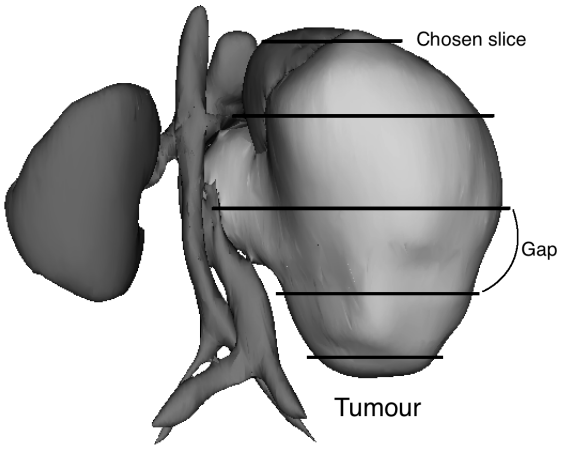

Figure 5: Notion of slice and gap in 3D tumoral segmentation

3.2.1. Vector optimization

To explore the possibility of ignoring certain slices, we thus introduced the notion of vector to express the data for the learning set LS. This allowed us to reduce our dataset. For each element of this vector, there are two possible values: 1 if the slice is used for training, 0 otherwise. The gap between each chosen slice is the same so as to recover information homogeneously at different levels and to succeed in segmenting all of the patient's other tumor slices with sufficient accuracy. For example, a gap of 2 means we take only 1 slice for the LS from every 3 slices, as shown in Table 1. Each gap produces several possible vectors (with a shift from one to other).

To do this, we use i vectors V_g, with a gap g, to feed the learning set LS. V_g is a part of V, the set of all possible combinations as:

$$V_g \in V = \{(\{0,1\}^n)\} \tag{8}$$

with n the number of slices to segment.

The learning set LS of the CNN contains all slices used to train the network and is defined as:

$$LS = \{S_j, ..., S_k\} \tag{9}$$

where $[S_j, S_k]$ are the slices of the tumor selected by the vector V_g.

With each gap producing several possible vectors, V_g is,in fact, a set of vectors. The following notation, $V_{g.h}$, is used for vector number h with a gap g such as:

$$V_g = \bigcup_{i=1}^{h} V_{g.i} \tag{10}$$

An example of the three possible vectors with a gap of 2 is presented in Table 1. Testing all these vectors for a given gap allows evaluation not only of the gap in the results, but also the impact of the slice positions.

	S_1	S_2	S_3	S_4	S_5	S_6	...	S_{60}	S_{61}
$V_{2.1}$	1	0	0	1	0	0	...	0	1
$V_{2.2}$	0	1	0	0	1	0	...	0	0
$V_{2.3}$	0	0	1	0	0	1	...	1	0

Table 1: Example of V_2 for $n = 61$

According to $V_{g.h}$, LS contains a restrained set of slices. We can build LS by applying a simple mathematical function f which selects the relevant slices following the values in $V_g.h$:

$$f : LS \times \{0, 1\} \rightarrow LS | \forall i \in \{1, ..., n\} \begin{cases} f(S_i, 0) = \emptyset \\ f(S_i, 1) = S_i \end{cases} \tag{11}$$

We also introduced a function F, using f and defining our final LS according to our vector:

$$F : V \rightarrow LS | F(V_{g.h}) = \bigcup_{i=1}^{n} \{f(S_i, V_{g.h_{i-1}})\} \tag{12}$$

3.2.2. Cardinality

The cardinality of the vectors set is $Card_V = 2^n$ but as we introduce the notion of gap, not all combinations are tested. The cardinality is then related to the chosen gap. The cardinality for a gap $Card_{V_g}$ is then:

$$\forall g \in \{0, ..., n-2\} \quad Card_{V_g} = g + 1 \tag{13}$$

since we have a gap of g between two included data, the set of possible vectors with this gap is g plus the first vector.

The cardinality for all the gaps we test $Card_{V_{test}}$ is :

$$Card_{V_{test}} = \sum_{g=0}^{p} (Card_{V_g}) \tag{14}$$

4. Results

4.1. Coefficients for evaluation of results

We used two scientific indicators in order to compare the results. The Dice and the IU. The Dice is commonly used by experts in medical imaging, and the IU is an index commonly used by the Imaging community.

The Dice coefficient, also known as the Sorensen coefficient, gives a similarity value (on $[0, 1]$) between two sets X and Y. In our case, X represents the pixels of the ground truth (i.e. the desired segmentation) and Y represents the pixels of the calculated segmentation given by our system. The above formula gives the details to calculate the value of the coefficient's value :

$$Dice = \frac{2 * |X \cap Y|}{|X| + |Y|}$$

The IU (Intersection over Union, also called Jaccard's index) is the mean of the IU_i of all the regions; the IU_i of region i is given by the above equation:

$$IU_i = \frac{|X \cap Y|}{|X \cup Y|} = \frac{n_{ii}}{n_{ii} + n_{ji} + n_{ij}}, \text{ where:}$$

- n_{ii} is the number of pixels correctly placed in region i,
- n_{ji} is the number of pixels that should have been in region i but are not,
- n_{ij} is the number of pixels that are in region i but should not be.

4.1.1. Training

The CNN we use for segmentation is the FCN-8s, for its effectiveness and smaller number of layers compared to other CNNs. We also use the same learning rate (1^{-12}), momentum (0.99) and weight decay (0.0005) as well as these pre-trained parameters of the conv1 to fc7 layer of the base PASCAL VOC 2012 Everingham et al. (2015). The last convolution is modified with two channel dimensions to predict two scores, the background and the tumor. The network is then trained with pre-trained parameters on a maximum training set of 61 slices, according to the chosen gap, from a single patient. To evaluate CNN performance, we used as a test set only slices which are not part of the training set, except for the particular case of $g = 0$. For this case, all slices were included in the training set so all of them had to be used.

We trained the networks using a Tesla Kepler K40 GPUs from Mésocentre at the University of Franche-Comté. Several vectors were tested with different gaps, from 0 to 10, and for each gap, we checked all possibilities according to $Card_{V_{test}}$. We thus had 65 possibilities of vectors, each vector having a size of 61 elements determining the slices to be used for training. For the vector with a gap of 0, all sets of slices were chosen for training.

4.2. Kidney and tumor segmentation

First, we need to evaluate results from our CBR and CNN systems in order to determine what approach is the best to segment both structures: kidney and tumor. The CBR system has been evaluated with a case base of 10 images from a single patient and a cross validation process versus FCN-8s with a training set of 12 random slices and a test set of 6 random slices from a single patient.

	CBR-RG		FCN-8s	
	Mean	Mean IU	Mean	Mean IU
Tumor	0.92	0.86	0.94	0.89
Kidney	0.83	0.71	0.36	0.23

Table 2: Dice and IU for kidney and tumor segmentation with CBR-region growing and CNN

Table 2 shows our preliminary results with a Dice and IU calculation. FCN-8s has the best scores for tumor segmentation with a mean DICE of 0.94, whereas the Region-Growing method enhanced with CBR (CBR-RG) was better for kidney segmentation. CNN is a very powerful technique but needs a large data set to express its potential. Moreover, neural networks can more easily segment objects with a regular and simple form. In our application, a kidney is a smaller structure than a tumor and may be badly deformed due to nephroplastoma. IThis explains the poor values for kidney segmentation obtained with FCN-8s. Concerning the tumor, both approaches have a close performance with a Dice of 0.92 for CBR and of 0.94 for FCN, so CNN has a small advantage. Another benefit is that CNN training is easier than building the Case Base. Moreover, CBR-RG is a sensitive method to leak phenomena which can lead to difficult segmentations. From these results, we can conclude CNN appears to be the best solution for tumor segmentation while CBR (with RG) is clearly more apt for the kidney.

4.3. Evaluation of the CBR system with adaptation phase

4.3.1. Evaluation of similarity metrics

Table 3 shows the different similarity values according to the formulas mainly used in image processing with each source case. The last column shows the retrieved case for each target case and each similarity measure. The retrieved case is the one with the higher similarity value, excluding the target case (a case cannot retrieve itself). Results are equivalent for both and, for each case the retrieved case is the same. There are only two differences. On the one hand,

Perner's formula appears more discriminative, with a higher difference between the similarities from one image to another. On the other hand, MSSIM appears globally more accurate because the similarity of two identical cases is equal to 1. Consequently, the MSSIM similarity was chosen.

4.3.2. System adaptation

In order to illustrate the effects of this adaptation algorithm, Figure 6 shows a comparison between an execution of our program, with and without including the adaptation phase. The kidney has a light grey label and the tumor a dark grey one. From left to right we have a CT-image, segmentation without an adaptation phase and finally segmentation with an adaptation phase. For the segmentation without an adaptation process, the seed was placed outside the top part of the kidney. Consequently, the top part was not correctly segmented and the result became irrelevant. Our adaptation phase avoids this situation and improves segmentation precision.

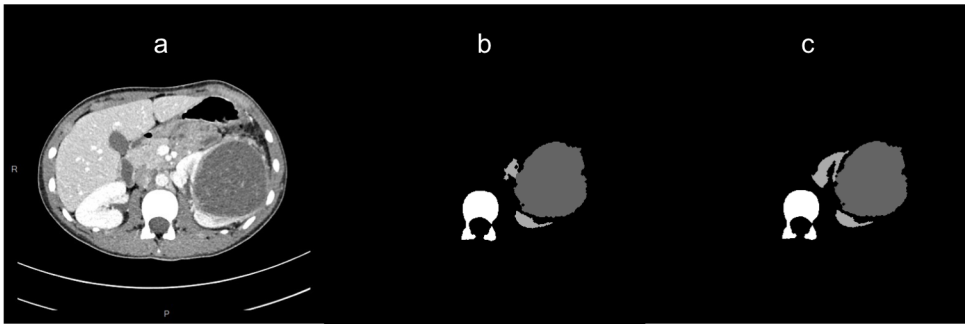

Figure 6: Comparison between an execution of our system with and without the adaptation phase: (a) CT-Scan, (b) segmentation without adaptation and (c) segmentation with adaptation

Without adaptation, the seed is only placed in the right position in 60% of our tested cases. Thanks to the adaptation process, seed placement was improved, attaining 100% in the right position.

Table 4 shows the results for kidney segmentation with DICE and IU. Using cross validation, the number corresponds to the tested scan as a target case,

		Case Base										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**retrieved case**
Case 1	s_1	0.999	0.788	0.537	0.299	0.155	0.908	0.733	0.458	0.227	0.100	6
	s_2	1.000	0.950	0.903	0.876	0.862	0.995	0.935	0.894	0.870	0.859	6
Case 2	s_1	0.788	0.999	0.749	0.511	0.368	0.696	0.945	0.670	0.440	0.312	7
	s_2	0.950	1.000	0.966	0.935	0.915	0.936	0.995	0.956	0.926	0.910	7
Case 3	s_1	0.537	0.749	0.999	0.761	0.618	0.446	0.805	0.920	0.690	0.562	8
	s_2	0.903	0.966	1.000	0.980	0.960	0.889	0.981	0.997	0.970	0.953	8
Case 4	s_1	0.299	0.512	0.761	0.999	0.857	0.207	0.566	0.841	0.929	0.801	9
	s_2	0.876	0.935	0.980	1.000	0.990	0.862	0.951	0.988	0.997	0.982	9
Case 5	s_1	0.156	0.368	0.618	0.857	0.999	0.064	0.423	0.698	0.928	0.944	10
	s_2	0.862	0.915	0.960	0.989	1.000	0.849	0.932	0.970	0.995	0.996	10
Case 6	s_1	0.908	0.696	0.446	0.207	0.064	0.999	0.641	0.366	0.136	0.008	1
	s_2	0.995	0.936	0.889	0.862	0.850	1.000	0.920	0.879	0.856	0.846	1
Case 7	s_1	0.733	0.945	0.805	0.566	0.423	0.641	0.999	0.725	0.495	0.367	2
	s_2	0.935	0.995	0.980	0.951	0,932	0.920	1.000	0.971	0.942	0.925	2
Case 8	s_1	0.458	0.670	0.920	0.841	0.698	0.366	0.725	0.999	0.770	0.642	3
	s_2	0.894	0.956	0.997	0.988	0.970	0.879	0.971	1.000	0.978	0.962	3
Case 9	s_1	0.227	0.440	0.690	0.929	0.928	0.136	0.495	0.770	0.999	0.872	4
	s_2	0.870	0.926	0.970	0.997	0.996	0.856	0.942	0.978	1.000	0.988	4
Case 10	s_1	0.100	0.312	0.562	0.801	0.944	0.008	0.367	0.642	0.872	0.999	5
	s_2	0.859	0.910	0.953	0.982	0.996	0.846	0.925	0.962	0.988	1.000	5

Table 3: Comparison of two similarity measures (MSSIM and Attig and Perner's formulas) for each case during cross validation

	Case	1	2	3	4	5	6
CBR RG	**Dice**	0.998	0.559	0.843	0.894	0.997	0.000
	IU	0.995	0.388	0.728	0.809	0.995	0.000
CBR RG + adaptation	**Dice**	0.998	0.999	0.993	0.894	0.997	0.998
	IU	0.995	0.999	0.987	0.809	0.995	0.997
	Case	**7**	**8**	**9**	**10**	**Mean**	**Median**
CBR RG	**Dice**	0.927	1.000	0.936	0.998	0.815	0.931
	IU	0.865	1.000	0.879	0.996	0.765	0.872
CBR RG + adaptation	**Dice**	0.927	1.000	0.973	0.998	0.978	0.997
	IU	0.865	1.000	0.948	0.996	0.959	0.995

Table 4: Dice and IU for kidney segmentation with CBR region growing with manual region growing as ground truth.

with the 9 others as source cases. Highlighted in grey are the cases presenting an improvement after the adaptation process. To check the efficiency of the adaptation process only, we compared it with the ground truth result of manual region growing. We highlighted in grey the cases presenting an improvement after the adaptation process. Most of the time, the CBR process succeeded in recovering the right coordinates to place the seed in the kidney. For these cases, adaptation did not improve the results because CBR did not need it. However, some segmentations (2, 3, 6 and 9) did not work well with the CBR process and seeds were placed at the wrong coordinates (outside the kidney). With the adaptation step, a clear improvement in results was observed (up to 99.8% for Case 6).

Table 5 the error rates for each segmented case during the retrieval phase. Errors were computed with respect to the threshold determined during manual region-growing segmentation and are a mean of all the seeds of the kidney (each seed has its own threshold and error rate). For most of them, we have a null error rate since the similarity between problem case and retrieved case is very high, therefore the thresholds are the same. For some cases, however, a higher error

	Retrived Case	Global error (%)	Local error (%)
Case 1	6	0.0	0.0
Case 2	7	0.0	0.0
Case 3	8	0.0	0.0
Case 4	9	20.9	17.5
Case 5	10	3.6	0.0
Case 6	1	0.0	0.0
Case 7	2	0.0	0.0
Case 8	3	0.0	0.0
Case 9	4	29.7	14.5
Case 10	5	0.0	0.0

Table 5: Mean errors for retrieved thresholds: local and global

rate is observed (cases 4 and 9), from 14.5 to 29.7%, which may explain why cases 4 and 9 don't have a high score in Table 4. This shows the main limitation of our CBR system without an adaptation phase for thresholds. This adaptation process represents an important perspective for improving its robustness. Case 7 has the lower result in spite of an error rate of 0%. This is because the thresholds used are closely related to the seed position. As the retrieved position is not the same as the seed position during manual region growing, it is logical for the scores to be different. Furthermore, the best parameters are not always determined manually. Thus, by coincidence, a CBR solution can be better and has a score lower than 1 due to differences.

4.3.3. Comparison with other approaches

In order to compare our CBR system with other classical approaches, we undertook a segmentation with a manual level set using Image J software and a CNN. For the CNN, we implemented FCN-8s architecture from Long et al., 2015 Long et al. (2015), trained with a training set of 10 slices (CT-Scans) corresponding to our case base and according to a cross validation strategy so as to have the same conditions as in CBR. Table 6 shows our results from these

	CBR		RG		Level Set		FCN-8s	
	Dice	IU	Dice	IU	Dice	IU	Dice	IU
Case 1	0.92	0.85	0.92	0.85	0.93	0.87	0.73	0.58
Case 2	0.89	0.80	0.88	0.79	0.88	0.79	0.78	0.64
Case 3	0.74	0.59	0.75	0.60	0.73	0.58	0.65	0.48
Case 4	0.74	0.59	0.79	0.65	0.75	0.60	0.52	0.35
Case 5	0.84	0.72	0.84	0.72	0.83	0.71	0.53	0.36
Case 6	0.95	0.90	0.95	0.90	0.90	0.83	0.47	0.31
Case 7	0.82	0.70	0.82	0.69	0.76	0.62	0.59	0.42
Case 8	0.76	0.62	0.76	0.62	0.18	0.10	0.60	0.42
Case 9	0.86	0.76	0.86	0.76	0.84	0.73	0.73	0.57
Case 10	0.76	0.61	0.77	0.63	0.73	0.58	0.35	0.21
Mean	**0.83**	**0.71**	**0.83**	**0.72**	**0.75**	**0.64**	**0.59**	**0.43**
Median	**0.83**	**0.71**	**0.83**	**0.70**	**0.79**	**0.66**	**0.59**	**0.42**

Table 6: Comparison of our CBR system and some other approaches. Dice and IU were calculated for segmentation of deformed kidney

different methods. The evaluated CBR system version is the one with Region Growing (RG) and an adaptation phase. As previously, evaluation is with a Dice and IU computation, but this time we computed them from true ground truth realized manually by pediatric surgeons.

The best results, presented in Table 6, were obtained with our CBR system. Indeed, we succeeded in obtaining results very close to manual RG with a mean Dice of 0.83. The Level Set technique shows good results for most of the cases, but it failed to segment Case 8 correctly. Overall, RG (manual or with CBR), performs more pertinent segmentations. CNN has the worst scores with a mean Dice of only 0.59, a logical result since CNNs are based on experiences and require a large database to yield interesting segmentations. CBR is also based on knowledge, but it requires much less data in order to work well.

4.4. Evaluation of the CNN trained following the OV²ASSION method

Tumor segmentation results are presented in Table 7. We evaluated the performance of our approach using segmentation accuracy, with Dice and Intersection over Union (IU) scores, between ground truth and predicted segmentation. The impact of gap and position were determined simultaneously. For each gap, we trained the network with slices not included in the training set only, except for the gap of 0 that we used as a reference and where both training and test sets are the same (all slices have been used to train it). The second column corresponds to subvectors which are all possible vectors with a gap g (gap is the same but there are different offsets). The first sub-vector begins with a value of 1 in the first position, the second with an offset of one and so forth. We computed a mean/median Dice and IU for each subvector as well as global scores for each gap. The last column shows whether the corresponding subvector contains both of the extremities (if the top slice and the last slice are in the training set).

4.4.1. Influence of gap

Table 7 shows values for gaps from 0 to 6, with the different vectors, but we evaluated as far as a gap of 10. As expected, the best results are obtained with a complete training set ($g = 0$) which corresponds to an overlearned situation, with a Dice of 0.926. After introduction of a variable gap, we see a decrease in scores as shown in Figure 7. To draw this mathematical curve, we computed mean Dice and mean IU given a particular gap value g (each gap having several possible vectors). We observed no clear distinction with a particular gap value. The decrease is gradual but we observe only a small difference between a gap of 2 and a gap of 3 (a Dice of approximately 0.90), whereas it corresponds to a 15% reduction in the training set. An optimal gap would thus be $g = 3$ which produces a mean Dice of 0.897 with only 26% of initial segmented slices. Yet, to determine the best gap value, we need to know what error rate is acceptable in our medical context with a computation of intersurgeon and intra-surgeon variance. Even with a gap of 10 (that is, with only 10% of slices from the patient

segmented manually), we obtained a Dice of 0.83.

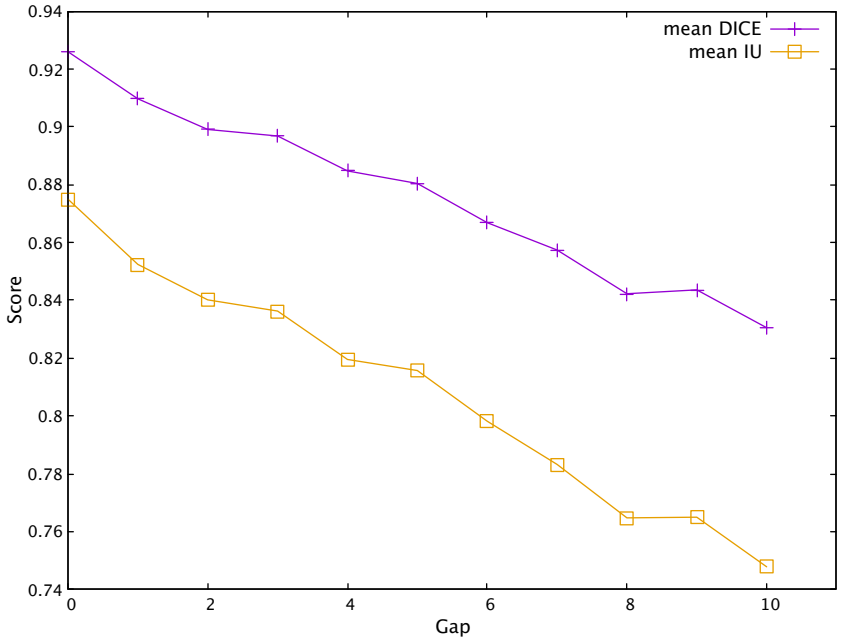

Figure 7: Evolution of mean Dice and mean IU following used gap

4.4.2. Influence of position

The different subvectors in Table 7 allow to determine the influence of slice position in the training set. For each gap value, we highlighted in gray the best subvectors (with higher values of Dice and IU). The last column shows that the best results are obtained once both extremities are included in the training set. These cases can be identified when both these conditions are validated:

$$\begin{cases} V_{g.h_0} = 1 \\ n \mod (g+1) = 1 \end{cases} \tag{15}$$

These results can be explained as tumor shape is more unpredictable at the extremities. From one slice to the next and looking back to previous ones, the shape may be very different, whereas the center of the tumor is more regular. Thus, inclusion of extremities increases scores. It also means we can include more slices in LS when we use extremities.

Gap	Vector	Mean Dice	Mean IU	Gap Dice 1	Gap IU 1	Median Dice	Median IU	Gap Dice 2	Gap IU 2	Both extremities
0	$\boldsymbol{V_0}$	0.926	0.875	0.926	0.875	0.955	0.914	0.955	0.914	Yes
1	$\boldsymbol{V_{1.1}}$	0.916	0.861	0.910	0.852	0.952	0.908	0.950	0.906	Yes
	$\boldsymbol{V_{1.2}}$	0.903	0.843			0.949	0.902			No
2	$\boldsymbol{V_{2.1}}$	0.914	0.856	0.899	0.840	0.947	0.899	0.945	0.896	Yes
	$\boldsymbol{V_{2.2}}$	0.896	0.832			0.937	0.882			No
	$\boldsymbol{V_{2.3}}$	0.887	0.832			0.951	0.906			No
3	$\boldsymbol{V_{3.1}}$	0.925	0.874	0.897	0.836	0.961	0.925	0.944	0.894	Yes
	$\boldsymbol{V_{3.2}}$	0.890	0.824			0.936	0.880			No
	$\boldsymbol{V_{3.3}}$	0.884	0.820			0.936	0.879			No
	$\boldsymbol{V_{3.4}}$	0.888	0.827			0.943	0.892			No
4	$\boldsymbol{V_{4.1}}$	0.920	0.867	0.885	0.816	0.949	0.903	0.936	0.880	Yes
	$\boldsymbol{V_{4.2}}$	0.881	0.810			0.930	0.870			No
	$\boldsymbol{V_{4.3}}$	0.887	0.821			0.941	0.888			No
	$\boldsymbol{V_{4.4}}$	0.882	0.817			0.937	0.881			No
	$\boldsymbol{V_{4.5}}$	0.853	0.783			0.924	0.859			No
5	$\boldsymbol{V_{5.1}}$	0.903	0.846	0.880	0.816	0.946	0.897	0.935	0.878	Yes
	$\boldsymbol{V_{5.2}}$	0.889	0.820			0.927	0.863			No
	$\boldsymbol{V_{5.3}}$	0.884	0.816			0.940	0.886			No
	$\boldsymbol{V_{5.4}}$	0.873	0.807			0.937	0.881			No
	$\boldsymbol{V_{5.5}}$	0.871	0.810			0.932	0.872			No
	$\boldsymbol{V_{5.6}}$	0.862	0.799			0.930	0.870			No

Table 7: Results of CNN according to the used gap and position of slices from training set for tumor segmentation. Gap Dice/IU 1 is the mean of mean Dice/IU for a given gap. Gap Dice/IU 2 is a mean of the median Dice/IU for a given gap.

5. Discussion

Our results prove that our CBR system can enhance and lead segmentations of organs and structures deformed by a singularity (the tumor). Indeed, using this technique, the deformed tumoral kidneys of our scans were widely retrieved and better segmented than with all the other tested methods. It mainly comes from our adaptation phase, to be sure that seeds are placed in the right position. Yet, this adaptation is based on a coherence of pixel intensity and, unfortunately, different structures can have close grey levels. Thus, it is possible for the adaptation to select the wrong pixel and to place seeds, for example, in the spine or ribs. Most of the time, the kidney is close enough to avoid this situation but it could theoretically occur. In this respect, it would be interesting to enhance our adaptation phase to prevent this possibility. One way of doing it would be to drive the pixel research in direction of the kidney, rather than visit all the neighbors around the initial seed. Another improvement would be to design another adaptation phase to adjust the thresholds recovered from the source case.

The OV^{2}ASSION method enables enhancement of CNN performance via an adaptation for a specific patient with a limited training set. We showed we can have efficient results with an overlearning on only 26% of slices from a single patient. In a medical context, where we can not have a large data set, it is an exciting way to improve segmentations. A surgeon could just segment 26% of the slices (according to a particular gap), feed the CNN and have a complete result. Ideally, we can imagine an automatic system which performs segmentation with an other approach, selects the best slices and feeds a CNN to have a complete segmented tumor. The main limitation is this reduced training set needs to be representative. That's why we did our experiments with a constant gap. In a real situation, an automatic system would select the best segmented slices but can not ensure this training set is really representative of the tumor.

Globally, RG produced better segmentations than the Level Set (LS) technique, though results are closed for most cases. Yet, for some cases, it is easier

and faster to parameter LS. This could lead to designing a CBR system using both approaches: a case might also include the segmentation technique (RG or LS), following the best way determined during the manual step. The CBR system could thus retrieve the segmentation approach used to solve a case and not only the parameters of a particular technique. In our approach, CNNs have poor results because of the lack of data (only 10 images). The main advantage of CBR with respect to CNN is that CBR is simultaneously an experience and a knowledge approach where CNN is just based on experience. This is why CBR can perform interesting results even with a small data set. We realize that we are in a particular situation where CNNs cannot give relevant results and it is sure that we could have better results with a larger training set, but it was our working conditions. CNNs are known as the better process for segmentation image in state of art. In consequences, we had to compare our system with this process in the same conditions. Moreover, CNNs have great difficulty segmenting such small structures as deformed kidneys without a large training set, but with larger structures such as tumors they can provide sufficiently accurate results. We can thus envisage a hybrid system with a CBR system for kidney segmentation and a CNN to segment tumors. Otherwise, CNN could be fed by a CBR system with a tumor segmentation on certain slices from the patient before complete segmentation by the neural network with OV^{2}ASSION.

In addition, for surgeons, kidney and tumor segmentation is insufficient. They also need to vizualize other structures such as blood vessels (arteries, veins, etc.) in order to evaluate the difficulty of the surgery. Yet, segmentation of these structures can be very difficult due to a lack of visibility. In some cases, vessels are not visible on CT-images. Because of their small size, placing a seed inside these objects is arduous. Only with a surgeon's experience can one draw vessels. Deep learning may be an attractive option. We can imagine overlearning on a particular slice and use CNN to segment the next slice before new training on both slices and so on. Furthermore, CBR can also theoretically integrate knowledge of the domain and thus be a good option in order to lead the search for such small structures.

6. Conclusion

We have designed a system based on CBR which can conduct the segmentation of tumoral kidneys and a method based on Deep learning which can perform accurate segmentation of nephroblastoma.

We have compared our CBR system to others approaches, both classical (commonly used in hospitals) and very recent (Deep learning). These highly promising results proved that we must look further into these approaches and that we are well on the way towards automatic segmentation of all organs. However, the small data set drives severe limitations for a generalization of these methods. A larger evaluation is needed to have a robsut system for image segmentation. The difficulty of our problem resides in the fact that tumoral kidneys are deformed by a totally unpredictable structure (the tumor). Nevertheless, our system can retrieve the positions of these organs.

Our method base on Deep Learning (OVÅSSION) consists of overlearning on a limited number of slices from a given patient to do a total segmentation during a second step. The notion of gap was introduced to retrieve the global tumor shape using a small training set. Results proved that segmentations performed following OVÅSSION are accurate. A gap of 3 allows a mean Dice very close to that with a gap of 0 despite that it represents a 74% reduction in training set size.

As regards these results, we hope to go further, both on CBR and Deep Learning. In a CBR-RG system, an adaptation phase for thresholds would be essential to improve its robustness and represents the next step. An upgrade of OVÅSSION is also an exciting perspective to allow better precision and its incorporation in a fully automatic system.

Acknowledgements

The authors thank the *European Community* (*European FEDER*) for financing this study by the *INTERREG V*, the *Communauté d'Agglomération du*

Grand Besançon, the *Cancéropôle Grand-Est* and Mesocenter of University of Bourgogne-Franche-Comté. The authors wish to thank Pr. Frederic Auber and Dr. Marion Lenoir-Auber of the Centre Hospitalier Régional Universitaire de Besançon for their expertise in nephroblastoma, and Loredane Vieille for performing the manual segmentations. The authors also wish to thank John Olsen for his help with the English language.

References

Attig, A., & Perner, P. (2009). A study on the case image description for learning the model of the watershed segmentation. *Transactions on Case-Based Reasoning*, *2*, 41–53.

Attig, A., & Perner, P. (2012). Incremental learning of the model for watershed-based image segmentation. In *Combinatorial Image Analaysis* (pp. 209–222). Springer. doi:10.1007/978-3-642-34732-0_16.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2015). Segnet: A deep convolutional encoder-decoder architecture for image segmentation.

Burgos-Artizzu, X. P., Ribeiro, A., Tellaeche, A., Pajares, G., & Fernández-Quintanilla, C. (2009). Improving weed pressure assessment using digital images from an experience-based reasoning approach. *Computers and electronics in agriculture*, *65*, 176–185. doi:10.1016/j.compag.2008.09.001.

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 424–432). Springer. doi:10.1007/978-3-319-46723-8_49.

Colliot, O., Camara, O., & Bloch, I. (2006). Integration of fuzzy spatial relations in deformable models—application to brain mri segmentation. *Pattern recognition*, *39*, 1401–1414. doi:10.1016/j.patcog.2006.02.022.

Cordier, A., Fuchs, B., & Mille, A. (2006). Engineering and learning of adaptation knowledge in case-based reasoning. In *EKAW* (pp. 303–317). Springer. doi:10.1007/11891451_27.

d'Aquin, M., Lieber, J., & Napoli, A. (2006). Adaptation knowledge acquisition: A case study for case-based decision support in oncology. *Computational Intelligence*, *22*, 161–176. doi:10.1111/j.1467-8640.2006.00281.x.

Diaz, F., Fdez-Riverola, F., & Corchado, J. M. (2006). gene-cbr: A case-based reasonig tool for cancer diagnosis using microarray data sets. *Computational Intelligence*, *22*, 254–268. doi:10.1111/j.1467-8640.2006.00287.x.

Dufour-Lussier, V., Le Ber, F., Lieber, J., & Martin, L. (2012). Adapting spatial and temporal cases. In *International Conference on Case-Based Reasoning* (pp. 77–91). Springer. doi:10.1007/978-3-642-32986-9_8.

Everingham, M., Eslami, S. A., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2015). The pascal visual object classes challenge: A retrospective. *International Journal of Computer Vision*, *111*, 98–136.

Frucci, M., Perner, P., & di Baja, G. S. (2008). Case-based reasoning for image segmentation by watershed transformation. In *Case-Based Reasoning on Images and Signals* (pp. 319–353). Springer. doi:10.1007/978-3-540-73180-1_11.

Fuchs, B., Lieber, J., Mille, A., & Napoli, A. (2000). An algorithm for adaptation in case-based reasoning. In *Proceedings of the 14th European Conference on Artificial Intelligence* (pp. 45–49). IOS Press.

Golobardes, E., Llora, X., Salamó, M., & Martı, J. (2002). Computer aided diagnosis with case-based reasoning and genetic algorithms. *Knowledge-Based Systems*, *15*, 45–52. doi:10.1016/S0950-7051(01)00120-4.

Henriet, J., & Chatonnay, P. (2013). Introduction of a combination vector to optimise the interpolation of numerical phantoms. *Expert Systems with Applications*, *40*, 492–499. doi:10.1016/j.eswa.2012.07.077.

Henriet, J., Chatonnay, P., & Leni, P.-E. (2014a). An iterative precision vector to optimise the cbr adaptation of equivox. *Engineering Applications of Artificial Intelligence*, *35*, 158–163. doi:10.1016/j.engappai.2014.06.017.

Henriet, J., & Lang, C. (2014). Introduction of a multiagent paradigm to optimize a case-based reasoning system designed to personalize three-dimensional numerical representations of human organs. *Biomedical Engineering: Applications, Basis and Communications*, *26*, 1450060. doi:10.4015/S1016237214500604.

Henriet, J., Leni, P.-E., Laurent, R., & Salomon, M. (2014b). Case-based reasoning adaptation of numerical representations of human organs by interpolation. *Expert Systems with Applications*, *41*, 260–266. doi:10.1016/j.eswa.2013.05.064.

Hong, S., Noh, H., & Han, B. (2015). Decoupled deep neural network for semi-supervised semantic segmentation. In *Advances in neural information processing systems* (pp. 1495–1503).

Hudelot, C., Atif, J., & Bloch, I. (2008). Fuzzy spatial relation ontology for image interpretation. *Fuzzy Sets and Systems*, *159*, 1929–1951.

Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, *36*, 61–78. doi:10.1016/j.media.2016.10.004.

Kato, Z., Zerubia, J. et al. (2012). Markov random fields in image segmentation. *Foundations and Trends® in Signal Processing*, *5*, 1–155. doi:10.1561/2000000035.

Kolodner, J. (2014). *Case-based reasoning*. Morgan Kaufmann.

Lieber, J. (2007). Application of the revision theory to adaptation in case-based reasoning: The conservative adaptation. *Case-Based Reasoning Research and Development*, (pp. 239–253). doi:10.1007/978-3-540-74141-1_17.

Lim, K. Y., & Mandava, R. (2018). A multi-phase semi-automatic approach for multisequence brain tumor image segmentation. *Expert Systems with Applications*, *112*, 288–300. doi:10.1016/j.eswa.2018.06.041.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Med Image Anal*, *42*, 66–88. doi:10.1016/j.media.2017.07.005.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).

Marling, C., Montani, S., Bichindaritz, I., & Funk, P. (2014). Synergistic case-based reasoning in medical domains. *Expert systems with applications*, *41*, 249–259. doi:10.1016/j.eswa.2013.05.063.

Melis, E., Lieber, J., & Napoli, A. (1998). Reformulation in case-based reasoning. *Advances in Case-Based Reasoning*, (pp. 172–183). doi:10.1007/BFb0056331.

Montani, S. (2009). Case-based reasoning for managing noncompliance with clinical guidelines. *Computational Intelligence*, *25*, 196–213. doi:10.1111/j.1467-8640.2009.00338.x.

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520–1528).

Panigrahi, L., Verma, K., & Singh, B. K. (2018). Ultrasound image segmentation using a novel multi-scale gaussian kernel fuzzy clustering and multi-scale vector field convolution. *Expert Systems with Applications*, *115*, 486–498. doi:10.1016/j.eswa.2018.08.013.

Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). Enet: A deep neural network architecture for real-time semantic segmentation.

Perner, P. (1999). An architecture for a cbr image segmentation system. *Engineering Applications of Artificial Intelligence*, *12*, 749–759. doi:10.1016/S0952-1976(99)00038-X.

Perner, P. (2001). Why case-based reasoning is attractive for image interpretation. In *Case-based reasoning research and development* (pp. 27–43). Springer. doi:10.1007/3-540-44593-5_3.

Perner, P., & Attig, A. (2011). Using prototype-based classification for automatic knowledge acquisition. In *Pattern Recognition, Machine Intelligence and Biometrics* (pp. 197–212). Springer. doi:10.1007/978-3-642-22407-2_8.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241). Springer. doi:10.1007/978-3-319-24574-4_28.

Thong, W., Kadoury, S., Piché, N., & Pal, C. J. (2016). Convolutional networks for kidney segmentation in contrast-enhanced ct scans. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, (pp. 1–6). doi:10.1080/21681163.2016.1148636.

Trzupek, M., Ogiela, M. R., & Tadeusiewicz, R. (2011). Intelligent image content semantic description for cardiac 3d visualisations. *Engineering Applications of Artificial Intelligence*, *24*, 1410–1418. doi:10.1016/j.engappai.2011.05.005.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*, 600–612.

Xia, X., & Kulis, B. (2017). W-net: A deep model for fully unsupervised image segmentation.

Zhou, X., Takayama, R., Wang, S., Zhou, X., Hara, T., & Fujita, H. (2017). Automated segmentation of 3d anatomical structures on ct images by using a deep convolutional network based on end-to-end learning approach. In *Medical Imaging 2017: Image Processing* (p. 1013324). International Society for Optics and Photonics volume 10133. doi:10.1117/12.2254201.