BGrowth: an efficient approach for the segmentation of vertebral compression fractures in magnetic resonance imaging

Jonathan S. Ramos[+1], Carolina Y. V. Watanabe[†],
Marcello H. Nogueira-Barbosa[⋆] and Agma J. M. Traina[+]

[+]Institute of Mathematics and Computer Science (ICMC), University of São Paulo (USP).
[†]Computer Science Department (DCC), University of Rondônia (UNIR)
[⋆]Ribeirão Preto Medical School (FMRP), University of São Paulo (USP).

Abstract

Segmentation of medical images is a critical issue: several process of analysis and classification rely on this segmentation. With the growing number of people presenting back pain and problems related to it, the automatic or semi-automatic segmentation of fractured vertebral bodies became a challenging task. In general, those fractures present several regions with non-homogeneous intensities and the dark regions are quite similar to the structures nearby. Aimed at overriding this challenge, in this paper we present a semi-automatic segmentation method, called Balanced Growth (BGrowth). The experimental results on a dataset with 102 crushed and 89 normal vertebrae show that our approach significantly outperforms well-known methods from the literature. We have achieved an accuracy up to 95% while keeping acceptable processing time performance, that is equivalent to the state-of-the-art methods. Moreover, BGrowth presents the best results even with a rough (sloppy) manual annotation (seed points).

Key-words: *Vertebral compression fractures, image segmentation, magnetic resonance imaging.*

1 Introduction

Spinal diseases are quite usual worldwide and can cause significant loss of function and quality of life (1). A very recurrent disease among older adults is the Vertebral Compression Fracture (VCF), which, in general, is caused by osteoporosis (benign) or bone metastasis (malignant) (2, 3, 4). In general, VCFs are early detected or diagnosed based on shape or texture using Magnetic Resonance Imaging (MRI) (5, 6, 3, 7). Usually, a specialist manually segments the Region of Interest (ROI) to aid the diagnose, which can be time consuming and prone to errors, due to inter and intra-subject variability and the subjective judgment that is employed (8).

However, it incorporates expert knowledge gained over several years. Figure 1 shows an example of manual segmentation over five lumbar vertebral bodies (L1-L5). Computer-Aided Diagnosis (CAD) systems with automatic or semi-automatic segmentation methods are quite useful in this case, as long as they present fast results and delineation closer/similar to the manual segmentation (8, 9). Hence, an accurate segmentation algorithm plays a significant role and can assist the medical specialist in surgical planning and evaluation of suitable treatments (10), for example, segmentation of Lung nodule (11, 12) and cortical bone (13, 14).

The semi or automatic segmentation of VCFs is a challenging task, due to non-homogeneous gray-scale intensities within the same vertebral body (for example, L5 in Figure 1). To overcome this issue, several one-seed-point approaches have been proposed. In (3) is presented the VBSeg method, which employs superpix-
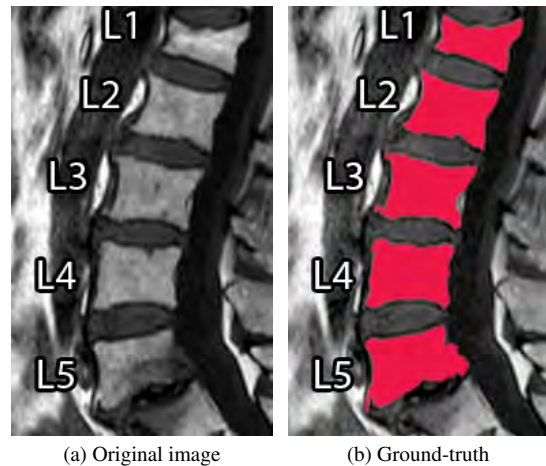

(a) Original image   (b) Ground-truth

Figure 1: Example of manual vertebral bodies segmentation.

els, region growing and Otsu threshold. Region growing techniques are used in (9), such as, snakes (Chan-Vese), Otsu and fuzzy c-means clustering, in order to compose a cooperative strategy for a dynamic ensemble of classification models. Although both works have presented very promising results, the VBSeg method presents a low segmentation performance on VCFs (61% and 74% Jaccard coefficient for malignant and benign VCFs, respectively). The cooperative strategy for classification still depends on the ground-truth.

In (8), the GrowCut (15) algorithm is employed for the segmentation of normal vertebral bodies. The Grow-Cut method and its faster version, named as Fast Grow-

Input: Image I and labels matrix L.
Output: Segmented binary image L == 1.
for $\forall(i,j)$ **do**   // **Step 1:** Set initial weights
    if $L(i,j) \neq 0$ **then**
        $W(i,j) \leftarrow 1.0$
    end
end
for $\forall(i,j)$ **do**   // **Step 2:** For every pixel and its
    for $\forall(i_n,j_n)$ **do**   // 8-Neighbours (i_n, j_n)
        // Will not expand unlabelled pixel
        if $L(i_n,j_n) \neq 0$ **then**
            $s \leftarrow W(i_n,j_n) \times \left[1 - \frac{|I(i,j)-I(i_n,j_n)|}{\max_{\forall i,j} I(i,j)} \right]$
            if $s > W(i,j)$ **then**
                $L(i,j) \leftarrow L(i_n,j_n)$
                $W(i,j) \leftarrow 0.5 \times W(i,j) + 0.5 \times s$
            end
        end
    end
end

Figure 2: Balanced Growth (BGrowth) method.

Cut (16) (presents slightly lower segmentation performance than the original GrowCut (16)), employ several seeds points inside and outside the object of interest and have been widely used for many medical MRI exams (especially in oncology) (17). However, to the best of our knowledge, GrowCut was not tested on VCFs.

Based on the formulation of segmentation as an energy minimization problem, many algorithms have been proposed, such as GrabCut (GB) and Lazy Snapping (LS). GrabCut uses Gibbs energy (18) and Gaussian Mixture Models (GMM) as soft segmentation for the background and foreground (19). LazySnapping works as an interactive image cutout tool and also uses Gibbs energy, combining graph cut with a pre-computed over-segmentation, e.g. Superpixels (20). Although GB and LS have not been used for the segmentation of VCFs, they have been adapted or used in several medical applications (21, 22, 23). We consider both approaches in our analysis to address a wider assessment of VCFs segmentation over the state-of-the-art methods.

Aimed at overriding the challenge of VCFs segmentation, we propose the Balanced Growth (BGrowth) method, which balances the weights employed during the regions expansion. The experimental results show that our approach significantly outperforms the methods from the literature, achieving 95% accuracy while keeping an processing time equivalent to the competitor methods. Moreover, we achieved the best results even with sloppy annotations.

The remainder of the paper is structured as follows. First, in section 2, we describe our proposed approach for the segmentation of vertebral bodies in MRI. Then, in section 3, we explore the materials and methods used in our work. Next, in section 4, we detail the experiment design, results and discussion. Finally, the conclusions are presented in section 5.

2 Balanced Growth: the proposed method

Our proposed approach is mainly based on the motto "unity is strength". We take into account that all pixels in the path of an expanding region have influence on its growth. We achieve "unity" by balancing the weights along the growing path, therefore the final segmentation is a "unanimity" among the pixels. This allows small intensities transitions (as occurs mostly often in fractured vertebral bodies) to be better delineated by **B**alanced **G**rowth (BGrowth).

For the sake of simplicity, we consider the segmentation of two regions (foreground and background) and gray-scale images. However, the method can be easily adapted to two or more regions and for any color space. Considering a digital image I and its annotations or labels regions as a matrix L, both with dimension $M \times N$ (M is the number of rows and N the number of columns), Figure 2 shows an overview of BGrowth. Each entry in L has value $-1, 0$ or 1, representing background, unlabelled and foreground, respectively.

The first step is to set up a weight matrix W as follows

$$W(i,j) = \begin{cases} 1.0 & \text{if } L(i,j) \neq 0 \\ 0.0 & \text{otherwise} \end{cases} \quad (1)$$

in which $i = 1, 2, ..., M$ and $j = 1, 2, ..., M$ are the matrices indexes and W has the same dimension as I and L.

The second step is to check each pixel and its 8-neighbours in the image. For each neighbouring pixel at index (i_n, j_n), a strength factor s is calculated as

$$s \leftarrow W(i_n, j_n) \times \left[1 - \frac{|I(i,j) - I(i_n, j_n)|}{\max_{\forall(i,j)} I(i,j)} \right], \quad (2)$$

in which the absolute pixel intensity difference is normalized by the maximum pixel value in the image and shifted by 1, which produces values within the range $[0, 1]$. Then, the strength factor is the product of the neighbouring weight and the pixel difference aforementioned. If s is greater than the current pixel weight, then the neighbouring pixel expands to the pixel at index (i, j): its label is changed and the new weight is balanced with the old one, i.e. $W(i,j) = 0.5 \times W(i,j) + 0.5 \times s$.

The balance induced by the weights produces smoother borders at regions with small pixel intensities. Moreover, as long as the interior annotation contains small portions of the dark and bright regions, it can work well even with rough interior and exterior seed points. To better illustrate this, Figure 3 shows a few iterations of BGrowth for a single vertebral body segmentation. Note that, at iteration 5, a few dark regions (outside of the vertebral body) are still part of the foreground and, as the balancing goes on, the foreground shrinks towards the ground-truth. The final result is quite close to the manual segmentation. Figure 4 shows another example of BGrowth's iterations on five lumbar vertebrae. Note that, the method works quite well even when a simple line is given as seed points. The only constraint is that the interior annotation has to comprise the dark/bright regions within the vertebral body.

As we work with the balance of old and new strength factors, it might happen that the process of expansion
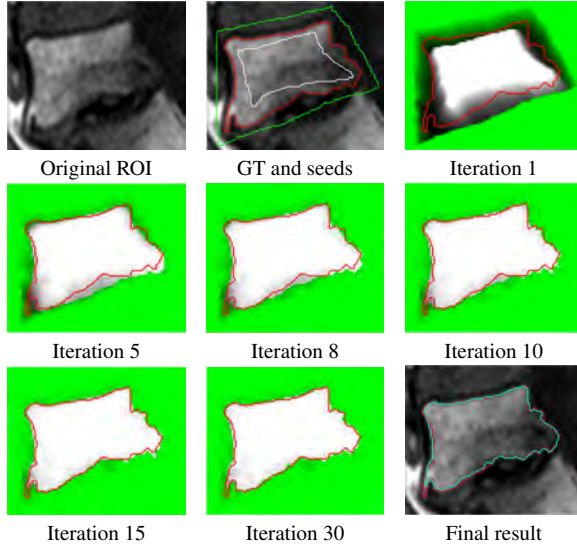
Figure 3: BGrowth' iterations: Ground-Truth (GT), interior annotation, exterior annotation and final result boundaries are outlined in red, white, green and cyan, respectively.

goes back and forth on the same pixel. Therefore, we would recommend the definition of a maximum number of iterations depending on the kind of image being segmented. We have empirically used a maximum of 30 iterations for the segmentation of a single vertebral body. For a whole exam, as exemplified in Figure 4, a few more iterations might be necessary.

In the worst case scenario, the algorithm has a time complexity of $M \times N \times 8 \times it = O(M \times N \times it)$, where it is the maximum number of iterations defined.

3 Materials and methods

In this section, we first describe the experimental dataset, which is composed of normal, malignant and benign vertebral bodies. Then, we present the comparison measures used. Next, we report the segmentation methods and parameters settings. Finally, the annotation scheme and computation set-up are explored.

3.1 Image dataset

Due to space limitations, we will show the results obtained from only one meaningful dataset of lumbar vertebral bodies (L1, L2, L3, L4, and L5), including 89 normal vertebral bodies and 102 with VCFs. Bone metastases occurred in 49 VCFs (malignant VCFs) and were confirmed by biopsy and histopathological analysis. The remaining 53 VCFs were benign fractures and, following clinical guidelines, not all of them were confirmed by biopsy or histopathological analysis. On these cases, a musculoskeletal radiologist with over 20 years of experience carefully reviewed the radiological information system (RIS) and the hospital information system (HIS) and supervised the manual segmentation (ground-truth). This study was approved by the Ethics Research Committee of the Ribeirão Preto Medical School - USP, where the dataset was acquired.
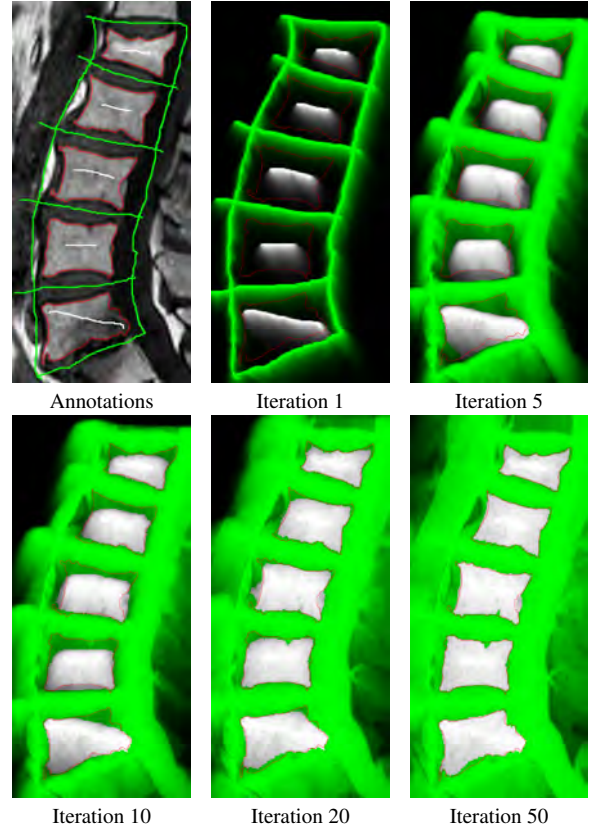

Figure 4: BGrowth' iterations on five lumbar: Ground-Truth, interior annotation and exterior annotation boundaries are outlined in red, white and green, respectively.

3.2 Comparison measures

We analyzed the Jaccard Coefficient J and Dice Score D (24, 25):

$$J(GT, Seg) = \frac{|GT \cap Seg|}{|GT \cup Seg|} \; ; \qquad (3)$$

$$D(GT, Seg) = \frac{2 \times |GT \cap Seg|}{|GT| + |Seg|} \; , \qquad (4)$$

in which GT represents the ground-truth region and Seg represents the region yielded by the segmentation technique.

For further analysis, we also employed the measures of accuracy (A) precision (P) and recall (R) (26, 27):

$$A = \frac{TP + TN}{TP + FP + TN + FN} \; ; \qquad (5)$$

$$P = \frac{TP}{TP + FP} \; ; \; R = \frac{TP}{TP + FN} \; , \qquad (6)$$

in which **True Positive (TP):** number of pixels correctly segmented as part of the vertebral body; **True Negative (TN):** number of pixels correctly segmented as part of the background; **False Positive (FP):** total of pixels miss-segmented as belonging to the vertebral body; **False Negative (FN):** number of pixels belonging to the vertebral body miss-segmented as part of the background.

The precision measures the percentage of pixels correctly segmented, considering the FPs. The higher the

Table 1: Summary of Acronyms/symbols used in this work.

Segmentation Methods		Comparison Measures	
Abbr.	Description	Symb.	Description
BG	Balanced Growth	A	Accuracy
CV	Chan-Vese	D	Dice-Score
GB	GrabCut	F	F-measure
GC	GrowCut	J	Jaccard coefficient
LS	LazySnapping	P	Precision
OT	Otsu	R	Recall

number of FPs, the lower the precision, indicating how much the method segments outside of he ground-truth. Likewise, the recall measures the percentage of pixels correctly segmented considering the FNs. The higher the number of FNs, the lower the recall, indicating how much is not segmented on the inside of the vertebral body.

Another measure that quantifies the balance between P and R is the F-measure, defined as ([28])

$$F = 2 \times \frac{P \times R}{P + R}. \tag{7}$$

[Table 1](#) shows a summary of the segmentation methods and comparison measures used in this work.

3.3 Segmentation algorithms

In order to evaluate the performance of BGrowth, we compared it with the methods[2] GrowCut (GC), Grab-Cut (GB), LazySnapping (LS), VBSeg (VBS), Snakes (Chan-Vese, CV) and Otsu threshold (OT).

3.4 Parameters and settings

The parameters for all methods that used superpixels were empirically set to $m \times n \times 0.25$ superpixels (25% of the total pixels in the image) for each ROI, in which m and n are the number of rows and columns of the ROI, respectively. The maximum number of iteration was set to 300 for Chan-Vese technique and 30 to GrowCut and BGrowth. The remainder of the parameters for all methods were set to default settings to avoid loss of generalizability. No pre or post-processing technique were applied to assure the same conditions for all segmentation methods.

3.5 Annotations

As [Figure 5](#) shows, the initial interior and exterior seeds annotation were performed in a "sloppy" way, i.e., no detailed boundary for accentuated curves were drawn. In general, the annotation looks like a rectangle.

3.6 Computational set-up

The experiments were performed on a desktop with a 3.60GHz Intel(R) Core(TM) i7 CPU and 16GB RAM using Matlab(R) version 2018a.

4 Experiments, results and discussion

The experiments are analyzed in three main parts. First we discuss results for the six measures aforementioned, then we analyzed their statistical difference for each segmentation method against Balanced Growth (BGrowth or BG). As a final analysis, we assess the
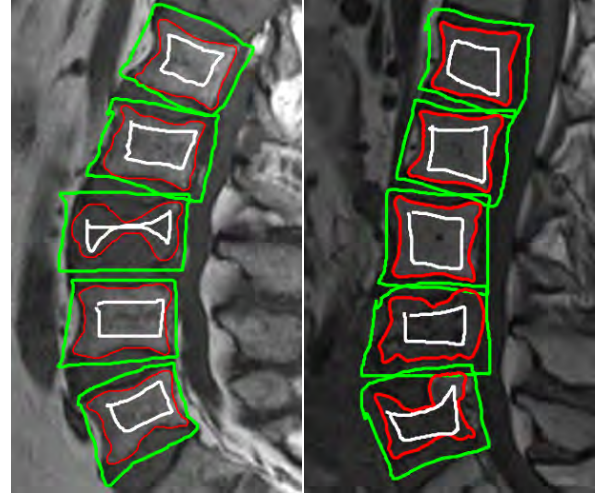
---

[2]Deep learning approaches were not used due to the small number of ROIs available.

Figure 5: Example of sloppy annotations (seed points): ground-truth (GT) in red; interior and exterior annotations in white and green, respectively.

impact of changing the percentage of manual annotation inside and outside each vertebral body for every approach that uses this kind of annotation.

4.1 Overall measures analysis

[Table 2](#) reports the average Accuracy (A), Jaccard coefficient (J) and Dice score (D) achieved by each segmentation method in our experiments separated by cases 'All', 'Normal', 'Benign' and 'Malignant'. Note that, for A, J and D:

- BGrowth (BG) presented the best results for all cases;

- LazySnapping (LS) presented the second best results for cases 'All', 'Normal' and 'Benign';

- GrowCut (GC) presented the second best results for 'Malignant' case;

- All the others methods presented results equal or below 80%;

[Table 3](#) reports the average Precision (P), Recall (R) and F-Measure (F) achieved by each method in our experiments separated by cases 'All', 'Normal', 'Benign' and 'Malignant'. Note that, for P, R and F:

- BGrowth (BG) presented the best balance between precision and recall (i.e. F-measure) and the second best results of precision and recall for all cases;

- Growcut (GC) presented the highest recall and the second best value of F-measure for the 'Malignant' case;

- LazySnapping (LS) presented the best results of precision for all cases and the second best values of F-measure for cases 'All', 'Normal' and 'Benign';

Although GrabCut (GB) presented the best values of recall for most cases, the precision is one of the lowest, which implicates that the method over-segments outside the region of the vertebral body. LazySnapping (LS)

Table 2: Comparison of Accuracy (A), Jaccard Coefficient (J) and Dice Score (D) for all methods considering distinct cases. The best value is in **bold** and the second one is underlined.

Method	Vertebral bodies considered (%)					
	All			Normal		
	A	J	D	A	J	D
BG	**95**±05	**85**±16	**91**±15	**96**±06	**88**±18	**92**±19
GC	93±05	81±15	88±15	94±06	85±18	90±18
LS	94±05	84±16	90±15	95±05	87±18	91±19
GB	76±12	58±16	72±16	78±11	62±17	75±17
CV	75±21	59±22	72±20	84±15	70±19	80±19
OT	78±13	53±26	64±28	84±09	66±21	76±22
VBS	71±110	–	–	–	–	–
	Benign			Malignant		
BG	**95**±05	**83**±15	**90**±14	**92**±02	**80**±07	**89**±04
GC	92±05	76±14	85±13	92±03	79±07	88±04
LS	94±05	82±15	89±14	92±04	79±09	88±06
GB	72±11	52±16	67±15	75±12	58±13	72±10
CV	76±22	57±23	69±21	55±18	41±12	57±12
OT	81±11	53±23	66±23	64±11	26±19	38±25
VBS	75±—	–	–	62±—	–	–

Table 3: Comparison of the Precision (P), Recall (R) and F-Measure (F) for all methods considering distinct cases. The best value is in **bold** and the second one is underlined.

Method	Vertebral bodies considered (%)					
	All			Normal		
	P	R	F	P	R	F
BG	88±16	94±14	**91**±15	91±19	93±19	**92**±19
GC	84±16	94±15	88±15	88±18	92±19	90±18
LS	**90**±17	91±15	90±15	**92**±19	90±19	91±19
GB	59±18	**96**±16	72±16	63±18	**95**±20	75±17
CV	66±24	88±17	72±20	75±21	90±16	80±19
OT	61±26	72±31	64±28	70±22	86±22	76±22
VBS	80±140	87±060	–	–	–	–
	Benign			Malignant		
BG	87±16	94±12	**90**±14	83±07	96±03	**89**±04
GC	79±15	94±13	85±13	82±08	**96**±02	88±04
LS	**89**±17	90±13	89±14	**84**±10	93±05	88±06
GB	52±17	**97**±14	67±15	60±17	96±07	72±10
CV	65±26	84±19	69±21	47±19	88±18	57±12
OT	63±22	72±25	66±23	39±26	42±31	38±25
VBS	85±—	86±—	–	68±—	88±—	–

presented the best precision, it means that the method segments a bigger region inside of the vertebral body. However, for the F-measure LS has not presented the best results.

Comparing our proposed BGrowth (BG) to GrowCut (GC), in general, BG keeps similar recall and achieves better precision, consequently, a higher F-measure. Figure 6 shows the segmentation results for BGrowth, GrowCut and LazySnapping in a case with three benign VCFs in L1, L3 and L5 (L2 and L4 are normal). Note that, the dark and bright regions within the same vertebral bodies difficult the segmentation for all methods, presenting spiked borders due to the neighbours structure intensities, which are very similar (dark). On these cases, even the manual segmentation is challenging and requires much experience of the human operator.

BG generated smother and more delineated borders, which are closer the ground-truth. The LS also produced promising results. However, sometimes it fails to delimit borders when the interior annotation is closer to the ground-truth and the exterior annotation is not, as shown in results for L4 in Figure 6.

Analyzing the running time, as Table 4 shows, BGrowth (BG) and GrowCut (GC) presented the same results. However, BG presented a smaller standard deviation, which indicates BG had less variation in the running time. Although the Otsu (OT) method presented the fastest running time, in most of the cases, OT presented the lowest values of accuracy (64% for the malignant case). Compared to BG and GC, LS is about five times slower, GB is almost four times slower and CV is highly slower (about 30 times).

Table 4: Running time comparison in seconds (s).

Method	Exam (5 ROIs)/s
BGrowth (BG)	0.154 ± 0.048
GrowCut (GC)	0.154 ± 0.066
LazySnapping (LS)	0.629 ± 0.301
GrabCut (GB)	0.492 ± 0.184
Chan-Vese (CV)	3.360 ± 0.360
Otsu (OT)	**0.007 ± 0.011**

4.2 Statistical testing

As the data presents several identical values, the Kolmogorov-Smirnov (29) test was used for the analysis of the normality of the data. The null hypothesis that the data follow a normal distribution was rejected for all measures on each segmentation algorithm. Therefore, the Wilcoxon (30) test was used to analyze if there are significant differences among the segmentation methods. In this test, the null hypothesis is that data from two measures, e.g. the precision from BGrowth (BG) and the precision from GrowCut (GC), are samples from continuous distributions with equal medians, against the alternative that they are not.

Table 5 shows the results for the Wilcoxon test. Note that, BGrowth (BG) presented significantly better results than GrabCut (GB), Chan-Vese (CV) and Otsu (OT) for each measure for all cases and than LazySnapping (LS) for the recall on every case. LS pre-

Table 5: Hypothesis testing results: the result ✓ indicates a rejection of the null hypothesis, i.e. there is significant difference among the same measures for the two methods, and × indicates a failure to reject the null hypothesis (no significant difference) at the 1% significance level.

BG ×	Vertebral bodies considered			
	All	Normal	Benign	Malignant
	A J D P R F	A J D P R F	A J D P R F	A J D P R F
GC	✓✓✓✓×✓	✓✓✓✓×✓	✓✓✓✓×✓	××××××
LS	×✓✓✓✓✓	×××✓✓×	××××✓×	××××✓×
GB	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓
CV	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓
OT	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓	✓✓✓✓✓✓

sented significantly better precision than BGrowth for cases 'All' and 'Normal'. On the other hand, on case

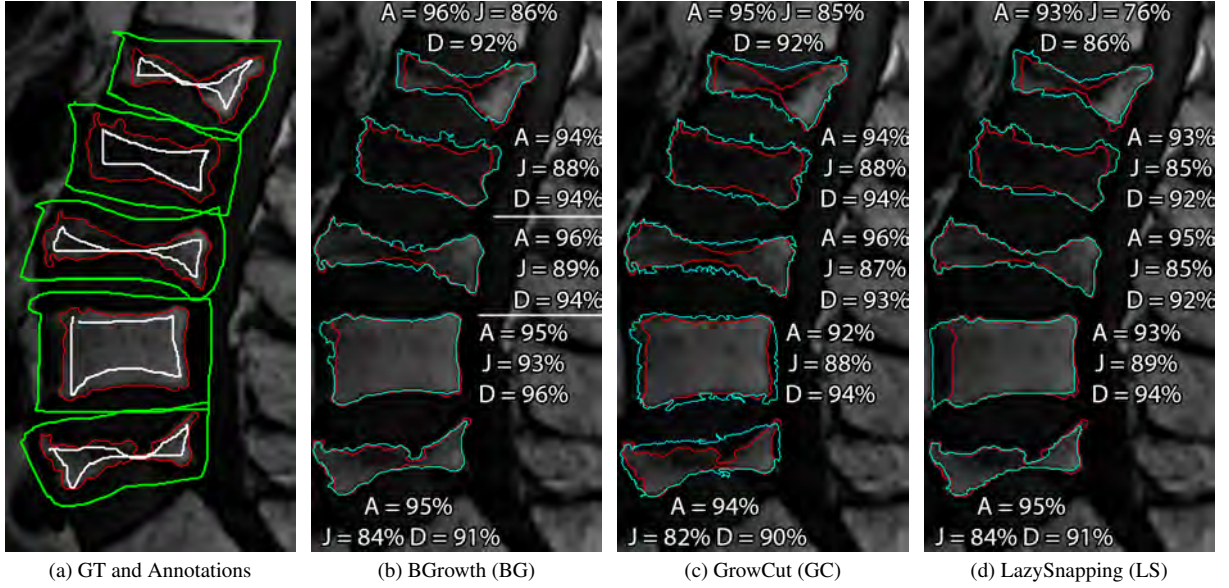|			
(a) GT and Annotations	(b) BGrowth (BG)	(c) GrowCut (GC)	(d) LazySnapping (LS)

Figure 6: Segmentation results for a case with all lumbar vertebrae with malignant compression fractures. The percentages shown represents the Jaccard Coefficient. Ground-truth in red and semi-automatic segmentation in cyan.

'All', BGrowth presented significantly better results than LazySnapping for Jaccard, Dice and F-measure. BGrowth improved GrowCut's performance in the three first cases ('All', 'Normal' and 'Benign'), for all measures (except the recall, in which BG kept performance similar to GC). On the case 'Malignant', BGrowth kept similar results to GrowCut for every measure.

4.3 Variation on the annotations

To assess the effect of the details employed on the interior and exterior manual annotations, we conducted two different analysis, which are detailed as follows:

1. We vary the percentage of interior annotation on each the vertebral body, and keep a sloppy exterior annotation done (as previous shown in Figure 5). For each vertebral body, a percentage of the ground-truth is extracted, always keeping at least a line that goes trough the dark and bright regions. The percentage starts at 10%, increasing by 10%, up to 100% (the ground-truth itself), summing up to 10 variations.

2. We vary the distance in pixels of the exterior annotation in relation to the ground-truth and keep a sloppy interior annotation. Starting at a distance of 3 pixels from the ground-truth, increasing by 3 up to 30 pixels, which sums up to 10 variations. If the distance surpass the image grid, the last pixel is assigned as boundary.

Figure 7 shows a few examples of variations of the interior percentages and the distances from the ground-truth used. Figure 8 shows the results for the interior and exterior variations for each one of the four cases using the Jaccard measure. Note that, for the interior annotation variation, BGrowth (BG) leads with the best results for all cases up to 60%. Besides, for all percentages and all cases, BG presented better results than GrowCut (GC). GrabCut (GB) presented the lowest results while
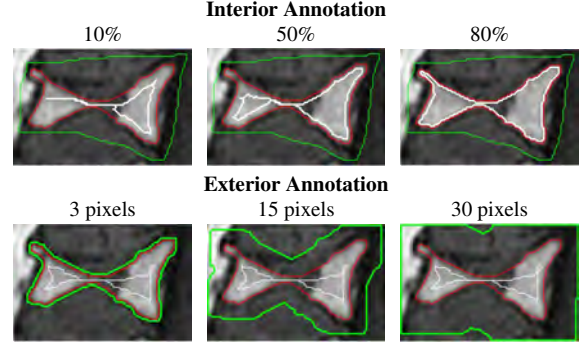

Figure 7: Examples of interior and exterior regions used.

LazySnapping (LS) presented the best results from 90% to 100%. However, in a real case scenario, 90% or 100% interior annotation rarely happens[3].

For the exterior annotation variation, GrabCut (GB) drops the performance really fast for all cases. GrowCut (GC) achieves its peak at 12 pixels for all cases, and then dropped the performance slowly. BGrowth (BG) presented better or similar performance than GC for all measures in all cases. LazySnapping (LS) presented results closer to BG with a higher difference from 6 to 18 pixels pixels for all cases. In general BG presented the best Jaccard values with "sloppy" annotation inside the vertebral body, which is very helpful, as the specialist does not need to spend much time drawing detailed curves at the interior region. The same holds for the exterior region annotation.

To further investigate, Figure 9 shows the results for the other five measures for the general case ('All'). Note that, for the interior annotation variation, BGrowth (BG) only presented lower results than GrowCut (GC) for the recall from 40% to 80% of interior annotation. How-

---

[3]It is practically the ground-truth itself, which takes too much time to annotate.
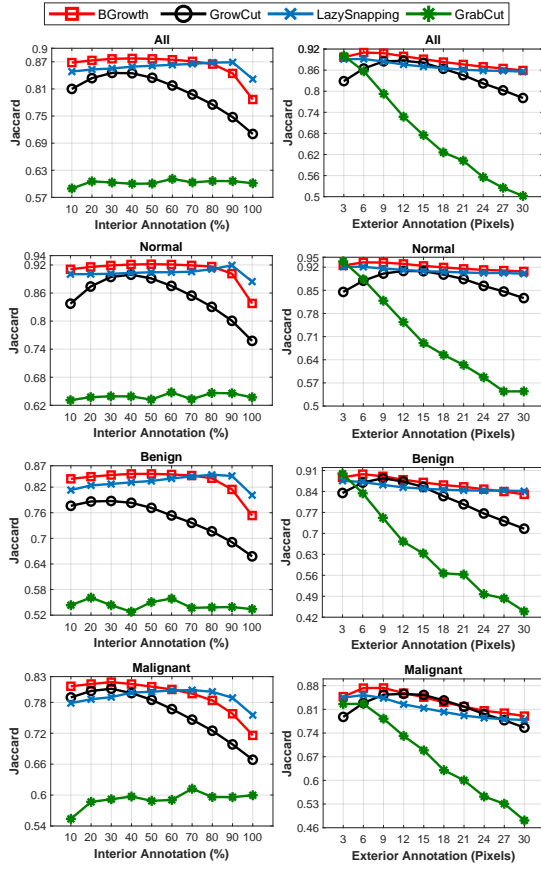
6

Figure 8: Comparison of Jaccard coefficient over annotation variations at the interior and exterior of each vertebral body.

Figure 9: Comparison of several measures over annotation variations at the interior and exterior of each vertebral body.

ever, the F-measure is higher for BG, which implicates that, in general, BG presented a better balance between precision and recall than GC. Although LazySnapping (LS) presented higher precision most of the time, its recall is one of the lowest. On the other hand, GrabCut (GB) presented one of the highest recall and lower values for the other measures.

Considering the interior annotation variation, for the accuracy, Dice, precision and F-measure, GB drops fast as the distance from the ground-truth boundary increases. GC, BG and LS presented quite similar behavior from 3 to 12 pixels. Over 12 pixels, there are small differences, in which GC presented the lowest results and BG and LS presented almost the same results.

In general, in a real case scenario, BGrowth (BG) is faster (as shown in Table 4) and produces better or similar results than LS even with sloppy interior and exterior annotations. Note that LS demands a more precise annotation in order to achieve better results, demanding more effort from the specialist.

5 Conclusions and future works

The semi-automatic segmentation of Vertebral Compression Fractures (VCFs) is a challenging task: in most cases, they present several regions with non-homogeneous intensities within the same vertebral body. We have investigated this challenge and we proposed an efficient and accurate method called BGrowth (BG), which balances the weights of the regions in ex-
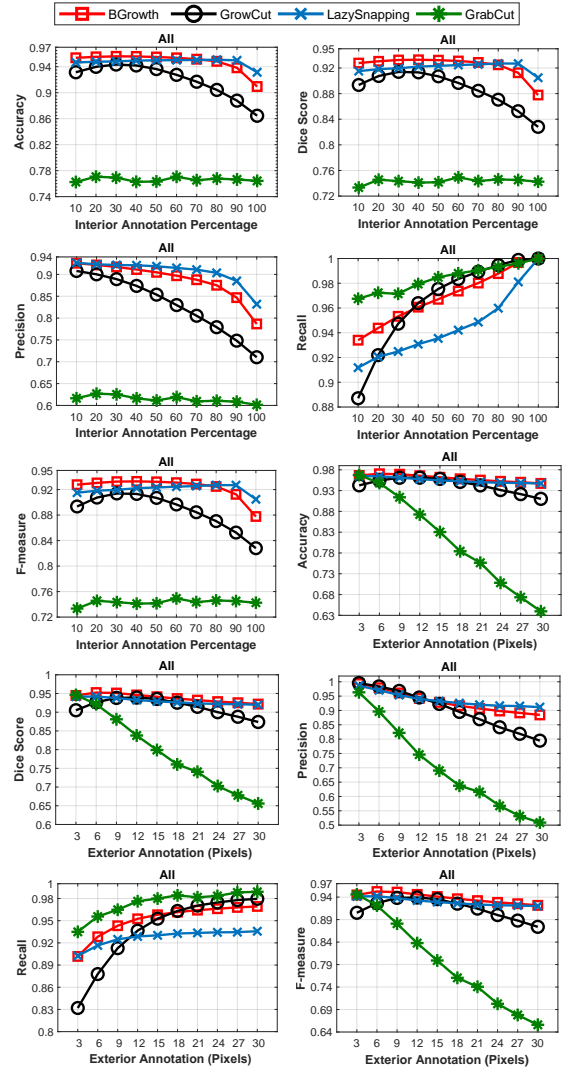
pansion. The segmentation performance obtained by BGrowth significantly outperforms other well-known methods from the literature and keeps an equivalent running time regarding the fastest competitors. Balanced Growth presents the best results with sloppy annotations, which demands less effort of the specialist on marking seeds points.

BGrowth presented an accuracy of 94% while Grow-Cut and LazySnapping presented 93% and 94%, respectively. Although LazySnapping presented results closer to BGrowth, its running time is almost five times slower.

As future work, we intend to improve BGrowth segmentation results by dynamically adapting the weights during the region expansion process and make it even easier for the specialist to annotate the seed points. Moreover, the segmentation results achieved can be used for feature extraction and classification.

Acknowledgments

References

1   MEGALE, R. Z. et al. Management of vertebral compression fracture in general practice: BEACH program. *PLoS ONE*, v. 12, n. 5, p. e0176351, 2017.

2   ZHAO, W. T. et al. Biomechanical effects of different vertebral heights after augmentation of osteoporotic vertebral compression fracture: a three-dimensional finite element analysis. *J Orthop Surg Res*, v. 13, n. 1, p. 32, Feb 2018.

3   BARBIERI, P. D. et al. Vertebral body segmentation of spine mr images using superpixels. In: JUNIOR, C. T. et al. (Ed.). *28th IEEE International Symposium on Computer-Based Medical Systems*. São Carlos and Ribeirão Preto, Brazil: Conference Publishing Services (CPS), 2015. p. 44–49. ISSN 1063-7125.

4   TEHRANZADEH, J.; TAO, C. Advances in mr imaging of vertebral collapse. *Seminars in Ultrasound, CT and MRI*, v. 25, n. 6, p. 440 – 460, 2004. ISSN 0887-2171. Imaging of Low Back Pain.

5   ANTANI, S.; LONG, L. R.; THOMA, G. R. Bridging the gap: Enabling cbir in medical applications. In: *2008 21st IEEE International Symposium on Computer-Based Medical Systems*. [S.l.: s.n.], 2008. p. 4–6. ISSN 1063-7125.

6   PAHOLPAK, P. et al. Evaluation of foraminal cross-sectional area in lumbar spondylolisthesis using kinematic mri. *European Journal of Orthopaedic Surgery & Traumatology*, Jul 2018. ISSN 1432-1068.

7   UETANI, M.; HASHMI, R.; HAYASHI, K. Malignant and benign compression fractures: differentiation and diagnostic pitfalls on mri. *Clinical Radiology*, v. 59, n. 2, p. 124 – 131, 2004. ISSN 0009-9260.

8   EGGER, J.; NIMSKY, C.; CHEN, X. Vertebral body segmentation with GrowCut: Initial experience, workflow and practical application. *SAGE Open Med*, v. 5, p. 2050312117740984, 2017.

9   CASTI, P. et al. Cooperative strategy for a dynamic ensemble of classification models in clinical applications: the case of MRI vertebral compression fractures. *International Journal of Computer Assisted Radiology and Surgery*, v. 12, n. 11, p. 1971–1983, Nov 2017.

10   DZENAN, Z. et al. Robust detection and segmentation for diagnosis of vertebral diseases using routine mr images. *Computer Graphics Forum*, v. 33, n. 6, p. 190–204, 2014.

11   EL-BAZ, A. et al. Computer-aided diagnosis systems for lung cancer: Challenges and methodologies. v. 2013, p. 942353, 2013.

12   SOLIMAN, A. et al. A novel cnn segmentation framework based on using new shape and appearance features. In: *2018 25th IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2018. p. 3488–3492. ISSN 2381-8549.

13   HAFRI, M. et al. Fuzzy energy based active contours model for hr-pqct cortical bone segmentation. In: *2016 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2016. p. 4334–4338. ISSN 2381-8549.

14   HAFRI, M. et al. Dual active contours model for hr-pqct cortical bone segmentation. In: *2016 23rd International Conference on Pattern Recognition (ICPR)*. [S.l.: s.n.], 2016. p. 2270–2275.

15   VEZHNEVETS, V.; KONOUCHINE, V. GrowCut - interactive multi-label N-D image segmentation by cellular automata. *International Conference on Computer Graphics and Vision - GraphiCon*, v. 1, Nov 2005.

16   ZHU, L. et al. An effective interactive medical image segmentation method using fast growcut. In: . [S.l.: s.n.], 2014. v. 17, n. WS.

17   JUNIOR, J. R. F. et al. Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. *Computer Methods and Programs in Biomedicine*, v. 159, p. 23 – 30, 2018. ISSN 0169-2607.

18   GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6, n. 6, p. 721–741, Nov 1984. ISSN 0162-8828.

19   ROTHER, C.; KOLMOGOROV, V.; BLAKE, A. GrabCut: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, ACM, New York, USA, v. 23, n. 3, p. 309–314, ago. 2004. ISSN 0730-0301.

20   LI, Y. et al. Lazy snapping. *ACM Trans. Graph.*, ACM, New York, USA, v. 23, n. 3, p. 303–308, ago. 2004. ISSN 0730-0301.

21   YONG, Z. et al. Grabcut image segmentation algorithm based on structure tensor. *The Journal of China Universities of Posts and Telecommunications*, v. 24, n. 2, p. 38 – 47, 2017. ISSN 1005-8885.

22   LU, Y.-W. et al. Segmentation method for medical image based on improved grabcut. *International Journal of Imaging Systems and Technology*, v. 27, n. 4, p. 383–390, 2017.

23   WU, S.; NAKAO, M.; MATSUDA, T. Automatic grabcut based lung extraction from endoscopic images with an initial boundary. In: *2016 IEEE 13th International Conference on Signal Processing (ICSP)*. [S.l.: s.n.], 2016. p. 1374–1378. ISSN 2164-5221.

24   JACCARD, P. The distribution of the flora in the alpine zone. *New Phytologist*, v. 11, n. 2, p. 37–50, fev. 1912.

25   SØRENSEN, T. *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. [S.l.]: I kommission hos E. Munksgaard, 1948. (Biologiske skrifter).

26   RAMOS, J. S. et al. How to speed up outliers removal in image matching. *Pattern Recognition Letters*, 2017. ISSN 0167-8655.

27   RAMOS, J. S.; WATANABE, C. Y. V.; TRAINA, A. J. M. Fomp: A novel preprocessing technique to speed-up the outlier removal from matched points. In: *2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. [S.l.: s.n.], 2016. p. 225–232.

28   GOUTTE, C.; GAUSSIER, E. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In: *Proceedings of the 27th European Conference on Advances in Information Retrieval Research*. Berlin, Heidelberg: [s.n.], 2005. (ECIR'05), p. 345–359. ISBN 3-540-25295-9, 978-3-540-25295-5.

29   MASSEY, F. J. The Kolmogorov-Smirnov test for goodness of fit. *Journal of the American Statistical Association*, American Statistical Association, v. 46, n. 253, p. 68–78, 1951.

30   WILCOXON, F.; KATTI, S.; WILCOX, R. Critical values and probability levels for the wilcoxon rank sum test and the wilcoxon signed rank test. *Selected Tables in Mathematical Statistics*, v. 1, p. 171–259, 1970.