Dynamically Balanced Online Random Forests for Interactive Scribble-based Segmentation

Guotai Wang[1], Maria A. Zuluaga[1], Rosalind Pratt[1,2], Michael Aertsen[3], Tom Doel[1], Maria Klusmann[4], Anna L. David[2], Jan Deprest[5], Tom Vercauteren[1], and Sébastien Ourselin[1]

[1]Translational Imaging Group, CMIC, University College London, UK
[2]Institute for Women's Health, University College London, UK
[3]Department of Radiology, University Hospitals KU Leuven, Belgium
[4]Department of Radiology, University College London Hospital, UK
[5]Department of Obstetrics, University Hospitals KU Leuven, Belgium

Abstract. Interactive scribble-and-learning-based segmentation is attractive for its good performance and reduced number of user interaction. Scribbles for foreground and background are often imbalanced. With the arrival of new scribbles, the imbalance ratio may change largely. Failing to deal with imbalanced training data and a changing imbalance ratio may lead to a decreased sensitivity and accuracy for segmentation. We propose a generic Dynamically Balanced Online Random Forest (DyBa ORF) to deal with these problems, with a combination of a dynamically balanced online Bagging method and a tree growing and shrinking strategy to update the random forests. We validated DyBa ORF on UCI machine learning data sets and applied it to two different clinical applications: 2D segmentation of the placenta from fetal MRI and adult lungs from radiographic images. Experiments show it outperforms traditional ORF in dealing with imbalanced data with a changing imbalance ratio, while maintaining a comparable accuracy and a higher efficiency compared with its offline counterpart. Our results demonstrate that DyBa ORF is more suitable than existing ORF for learning-based interactive image segmentation.

1 Introduction

Image segmentation plays an important role in a wide range of medical image processing and analysis tasks. Accurate and automatic segmentation remains difficult to achieve in many applications. Alternatively, user-interactive segmentation methods are widely used for higher accuracy and robustness. For example, TurtleSeg [9] uses initial user-provided contours in a few 2D slices for 3D segmentation. Graph Cuts [2] and Random Walks [5] learn a probability model from user-provided scribbles drawn on the foreground and background. To achieve a good interaction efficiency, machine learning methods [11, 6, 10] have been used to learn image features from the user inputs with reduced amount of interaction.

For interactive segmentation methods that learn from user-provided scribbles, several issues in the interaction and learning process should be considered.

Firstly, scribbles given at once are not always sufficient, and new scribbles can be added to rectify the segmentation results. To save additional training time when new scribbles arrive, online learning is preferred and is expected to achieve a correct rate of classification which is comparable with learning from scratch. Secondly, the amount of scribbles for the background and the foreground may be largely imbalanced. This could lead to a high error rate of classification for the minority class [4]. Furthermore, when more scribbles are added, the imbalance ratio between the foreground scribbles and background scribbles may change greatly. The learning algorithm should be adaptive to the changing imbalance ratio in order to deal with the imbalance problem correctly every time new scribbles are added.

In recent years, Random Forests [3] have been prevailingly used in computer vision and medical image analysis tasks due to their efficiency and high performance. For interactive segmentation [8], the sequential arrival of scribbles can be handled by Online Random Forests (ORF) [7]. However, the data imbalance problem was not explicitly addressed in that work . The ORF-based segmentation used by Barinova et al. [1] handles imbalanced data by re-sampling the training data with different sampling rates for different classes, but it assumes that the imbalance ratio of different classes does not change during online learning. The shortcoming is that when newly arrived training data leads to a large change of the imbalance ratio, the trees are not adapted to the new imbalance ratio. As a result, such forests may have a decreased accuracy of classification compared with their offline counterpart in interactive segmentation.

In this work, we propose a generic Dynamically Balanced Online Random Forest (DyBa ORF) to deal with incremental and imbalanced training data with a changing imbalance ratio. We applied DyBa ORF to two different applications: learning-based interactive segmentation of the placenta from fetal MRI and adult lungs from radiographs. In such applications, the segmentation task is challenging due to the low contrast between the target and background and inhomogeneous appearance of the target. This motivates us to use high level features combined with DyBa ORF-based learning rather than a traditional Gaussian Mixture Model, which is often used to model low dimensional features and not well suited to online learning. We investigated how DyBa ORF outperforms traditional ORF in these two applications, with its ability to achieve a comparable accuracy and higher efficiency compared with its offline counterpart.

2   Methods

Traditional Online Random Forests. A Random Forest [3] is a set of N binary decision trees with split nodes and leaf nodes. A split node executes a binary test to propagate a sample to its left or right child. A leaf node stores all the training samples that have been propagated to it and uses the distribution of class labels in that leaf for prediction. To overcome over-fitting, the training set of each tree is obtained by randomly re-sampling (a.k.a. Bagging) the original training set for the forest. To deal with online learning, the ORF [7] uses online

Bagging to model the sequential arrival of data as a Poisson distribution $\text{Pois}(\lambda)$ with a rate λ. Each tree is updated on each new training sample k times where $k \sim \text{Pois}(\lambda)$ and the expectation of k is λ. To deal with imbalanced data, Barinova et al. [1] used different λ for different classes based on the imbalance ratio. After receiving new data which leads to a new imbalance ratio (which may or may not be equal to the previous one), their method samples them with a rate based on the new imbalance ratio to grow existing trees, but does not update the set of existing sampled training data which has been sampled with a rate based on the old imbalance ratio. Thus, it fails to be truly adaptive to imbalance ratio changes.

Dynamically Balanced Online Bagging. For the sake of simplicity and without loss of generality, we focus on a binary classification problem. Suppose at an initial stage of online learning, the training data for the forests is represented by a tuple $\mathcal{S}_0(\mathcal{P}_0, \mathcal{N}_0)$ where $\mathcal{P}_0$ ($\mathcal{N}_0$) is a set of positive (negative) data. The initial imbalance ratio is defined as $\gamma_0 = |\mathcal{N}_0|/|\mathcal{P}_0|$. To deal with imbalanced data, we down-sample the majority class for efficiency [4]. Supposing $\text{Pois}(\lambda)$ is used to re-sample the minority class, $\text{Pois}(\lambda_{p0})$ and $\text{Pois}(\lambda_{n0})$ are used to re-sample $\mathcal{P}_0$ and $\mathcal{N}_0$ respectively:

$$\lambda_{p0} = \begin{cases} \lambda, & \text{if } \gamma_0 \geq 1.0 \\ \lambda\gamma_0, & \text{otherwise} \end{cases} ; \qquad \lambda_{n0} = \begin{cases} \lambda/\gamma_0, & \text{if } \gamma_0 \geq 1.0 \\ \lambda, & \text{otherwise} \end{cases} \qquad (1)$$

Thus each sample in $\mathcal{P}_0$ ($\mathcal{N}_0$) is expected to be sampled λ_{p0} (λ_{n0}) times. We denote the sampled training set for one tree as $\dot{\mathcal{S}}_0(\dot{\mathcal{P}}_0, \dot{\mathcal{N}}_0)$, where $\dot{\mathcal{P}}_0$ and $\dot{\mathcal{N}}_0$ are sampled from $\mathcal{P}_0$ and $\mathcal{N}_0$ respectively. $|\dot{\mathcal{P}}_0|$ has an expectation of $\lambda_{p0}|\mathcal{P}_0|$ and $|\dot{\mathcal{N}}_0|$ has an expectation of $\lambda_{n0}|\mathcal{N}_0| = \lambda_{n0}\gamma_0|\mathcal{P}_0| = \lambda_{p0}|\mathcal{P}_0|$. Thus, the sampled training data $\dot{\mathcal{S}}_0$ is balanced and used to generate the tree.

When a set of new training data $\mathcal{S}'(\mathcal{P}', \mathcal{N}')$ arrive, $\mathcal{S}'$ is added into $\mathcal{S}_0$. A merged training data set $\mathcal{S}_1(\mathcal{P}_1, \mathcal{N}_1)$ is obtained with a new imbalance ratio of $\gamma_1 = |\mathcal{N}_1|/|\mathcal{P}_1|$. In an offline situation, $\text{Pois}(\lambda_{p1})$ and $\text{Pois}(\lambda_{n1})$ should be used to sample $\mathcal{P}_1$ (obtain $\dot{\mathcal{P}}_1$) and $\mathcal{N}_1$ (obtain $\dot{\mathcal{N}}_1$) respectively, where λ_{p1} and λ_{n1} are defined based on γ_1 and λ in the same way as shown in Eq.(1). Instead of sampling $\mathcal{P}_1$ and $\mathcal{N}_1$ to get $\dot{\mathcal{P}}_1$ and $\dot{\mathcal{N}}_1$ from scratch, we dynamically update $\dot{\mathcal{P}}_0$ and $\dot{\mathcal{N}}_0$ to obtain $\dot{\mathcal{P}}_1$ and $\dot{\mathcal{N}}_1$: we generate an Add Set $\mathcal{A}$ and a Remove Set $\mathcal{R}$ from both $\mathcal{S}'$ and $\mathcal{S}_0$ based on the imbalance ratio change in the following way.

For $\mathcal{S}'$, a standard balanced sampling procedure is applied by using $\text{Pois}(\lambda_{p1})$ and $\text{Pois}(\lambda_{n1})$ to sample $\mathcal{P}'$ (obtain $\dot{\mathcal{P}}'$) and $\mathcal{N}'$ (obtain $\dot{\mathcal{N}}'$) respectively and get a sampled subset $\dot{\mathcal{S}}'(\dot{\mathcal{P}}', \dot{\mathcal{N}}')$. $\dot{\mathcal{S}}'$ is added to $\mathcal{A}$.

For $\mathcal{S}_0$, we deal with its positive subset $\mathcal{P}_0$ first. With the new Poisson distribution $\text{Pois}(\lambda_{p1})$, each sample in $\mathcal{P}_0$ is expected to be sampled λ_{p1} times. The expected difference of sampling rate between before and after $\mathcal{P}'$ arrives is $\delta_{p0} = \lambda_{p1} - \lambda_{p0}$. If $\delta_{p0} > 0$, it means after the new data arrives, more positive data should be sampled from $\mathcal{P}_0$ in order to keep the same sampling rate $\text{Pois}(\lambda_{p1})$ as used for $\mathcal{P}'$. We additionally sample $\mathcal{P}_0$ with $\text{Pois}(\delta_{p0})$ to get an

Add Set $\mathcal{A}_p$, and add it to $\mathcal{A}$. If $\delta_{p0} < 0$, it means after the new data arrives, less positive samples from $\dot{\mathcal{P}}_0$ are needed to keep the same sampling rate $\text{Pois}(\lambda_{p1})$ as used for $\mathcal{P}'$. We generate a random number $r \sim \text{Pois}(|\delta_{p0}| \times |\mathcal{P}_0|)$ and get $\min(r, |\dot{\mathcal{P}}_0|)$ samples from $\dot{\mathcal{P}}_0$. We denote these sampled data as a Remove Set $\mathcal{R}_p$, and then add $\mathcal{R}_p$ to $\mathcal{R}$.

The same steps are used to deal with $\mathcal{S}_0$'s negative subset $\mathcal{N}_0$, so that either an Add Set $\mathcal{A}_n$ or a Remove Set $\mathcal{R}_n$ is obtained. Thus we get the whole Add Set $\mathcal{A} = \dot{\mathcal{S}}' \cup \mathcal{A}_p \cup \mathcal{A}_n$ and the whole Remove Set $\mathcal{R} = \mathcal{R}_p \cup \mathcal{R}_n$. To get the updated training sample set $\dot{\mathcal{S}}_1$ for a tree on the fly, we remove $\mathcal{R}$ from $\dot{\mathcal{S}}_0$ and add $\mathcal{A}$ to it: $\dot{\mathcal{S}}_1 = (\dot{\mathcal{S}}_0 - \mathcal{R}) \cup \mathcal{A}$. Thanks to the way $\mathcal{R}$ and $\mathcal{A}$ are generated, $\dot{\mathcal{S}}_1$ is balanced and adapted to the new imbalance ratio γ_1.

Tree Growing and Shrinking. Instead of reconstructing trees from scratch, we use the Remove Set $\mathcal{R}$ and Add Set $\mathcal{A}$ to update an existing tree that has been constructed based on $\dot{\mathcal{S}}_0$, to make the updated tree adapted to the imbalance ratio change. Each sample in $\mathcal{R}$ and $\mathcal{A}$ is propagated from the root to a leaf. Assuming a sub set $\mathcal{R}_l$ of $\mathcal{R}$ and a sub set $\mathcal{A}_l$ of $\mathcal{A}$ fall into one certain leaf l with an existing sample set $\mathcal{S}_{l\_old}$, the sample set of l is updated as $\mathcal{S}_{l\_new} = (\mathcal{S}_{l\_old} - \mathcal{R}_l) \cup \mathcal{A}_l$. Then tree growing or shrinking is implemented on l based on $\mathcal{S}_{l\_new}$. If $|\mathcal{S}_{l\_new}| > 0$, a split test is executed for l and its children are created (growing) if applicable based on the same split rules as used in the tree constructing stage [3]. If $|\mathcal{S}_{l\_new}| = 0$, l is deleted (shrinking). Its parent merges the left and right child and becomes a leaf. The parent of a deleted leaf is tested for growing or shrinking again if applicable.

Applying DyBa ORF to Interactive Image Segmentation. In our image segmentation tasks, DyBa ORF learns from user-gradually-provided scribbles, and it predicts the probability each pixel being foreground. The features are extracted from a 9×9 region of interest (ROI) centered on each pixel [8]. We use gray level features based on the mean and standard deviation of intensity, histogram of oriented gradients (HOG), Haar wavelet, and texture features from gray level co-occurrence matrix (GLCM). The probability given by DyBa ORF is combined with a Conditional Random Field (CRF) [2] for a spatial regularization.

3 Experiments and Results

DyBa ORF was compared with three counterparts: 1) a traditional ORF [1] with multiple Poisson distributions based on Eq. (1) (MP ORF), 2) a traditional ORF with a single Poisson distribution $\text{Pois}(\lambda)$ (SP ORF), and 3) an offline counterpart (OffBa RF) which learns from scratch when new data arrives. The parameter settings were: $\lambda = 1.0$, tree number 50, maximal tree depth 20, minimal sample number for split 6. The code was implemented in C++[1].

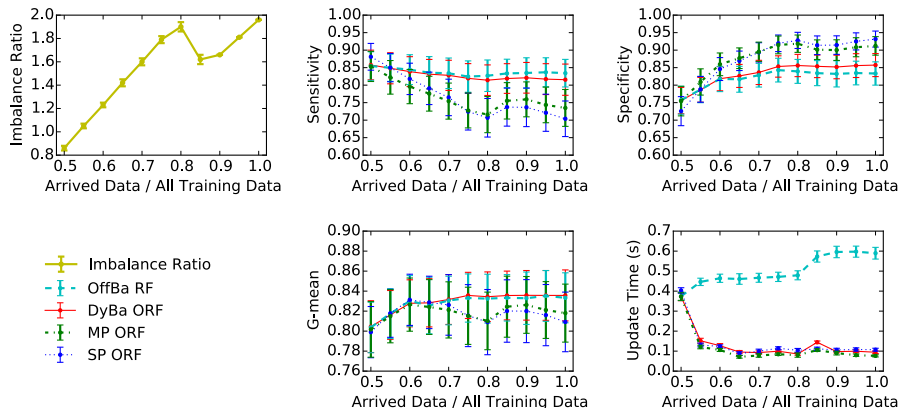[1] The code is available at https://github.com/gift-surg/DyBaORF

Fig. 1. Performance of DyBa ORF and counterparts on UCI QSAR biodegradation data set. Training data was gradually obtained from 50% to 100%.

Validation of DyBa ORF. Firstly, we validate DyBa ORF as an online learning algorithm with four of the UCI data sets[2] that are widely used: QSAR biodegradation, Musk (Version 1), Cardiotocography and Wine. The positive class labels for them are "RB","1","8" and "8" respectively. Each of these data sets has an imbalance between the positive and negative class. We used a Monte Carlo cross-validation with 100 repetition times. In each repetition, 20% positive samples and 20% negative samples were randomly selected to constitute test data. The remaining 80% samples were used as training data $\mathcal{T}$ in an online manner. The initial training set $\mathcal{S}_0$ contained the first 50% of $\mathcal{T}$ and it was gradually enlarged by the second 50% of $\mathcal{T}$, with 5% of $\mathcal{T}$ arriving each time in the same order as the appeared in $\mathcal{T}$.

Table 1. Comparison of G-mean on four UCI data sets after 100% training data arrived in online learning. The bold font shows values that are not significantly different from the corresponding results of OffBa RF(p-value>0.05). The G-mean of SP ORF on Wine is zero due to classifying all the samples into the negative class.

Data Set	Imbalance Ratio γ	G-mean (%)			
		OffBa RF	DyBa ORF	MP ORF	SP ORF
Biodegradation	1.23±0.47	83.33±2.50	**83.57±2.55**	81.80±2.90	80.92±2.98
Musk (version 1)	0.72±0.36	82.78±4.44	**83.06±4.08**	73.83±6.95	81.65±4.93
Cardiotocography	19.16±3.46	97.06±1.69	**97.09±1.56**	95.52±1.16	87.59±5.41
Wine	27.42±2.90	76.14±3.34	**76.50±3.91**	74.99±4.39	0.00±0.00

We measured the update time when new data arrived, sensitivity, specificity, and G-mean which is defined by G-mean=$\sqrt{\text{sensitivity} \times \text{specificity}}$. Ta-

---

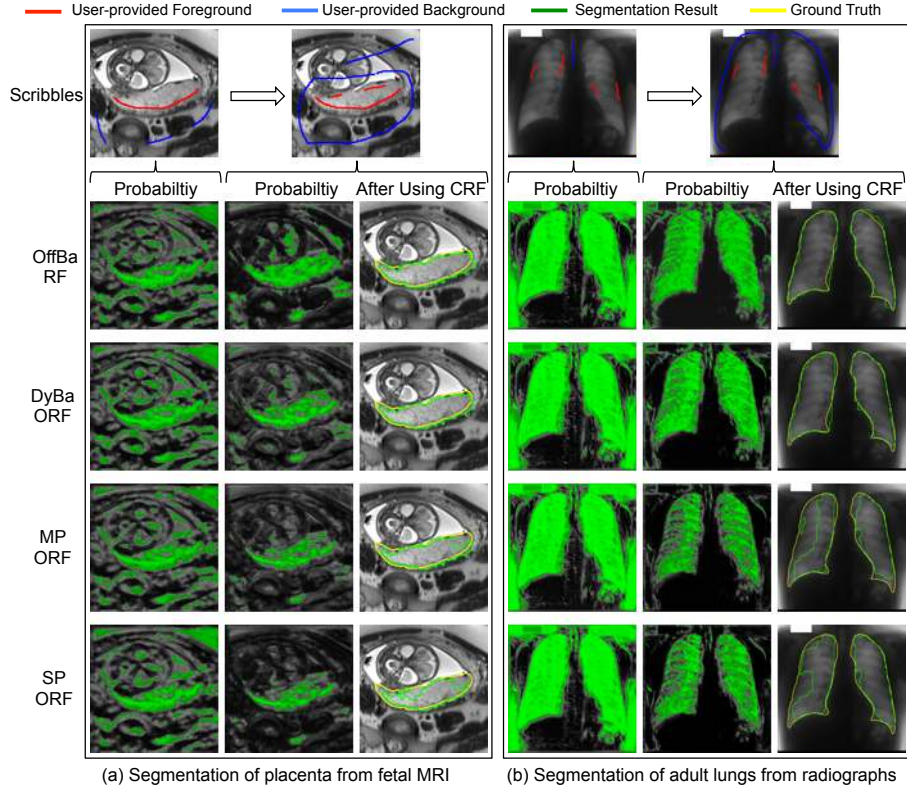[2] http://archive.ics.uci.edu/ml/datasets.html

Fig. 2. Visual comparison of DyBa ORF and counterparts in segmentation of (a) placenta from fetal MRI and (b) adult lungs from radiographs. The first row in each sub-figure shows two stages of interaction, where scribbles are extended with changing imbalance ratio. Probability higher than 0.5 is highlighted by green color. The last column in (a) and (b) show the final segmentation and the ground truth.

ble 1 shows the final G-mean on all the four datasets after 100% $\mathcal{T}$ arrived. The performances on the QSAR biodegradation data set are presented in Fig. 1, which shows a decreasing sensitivity and increasing specificity for SP ORF and MP ORF. In contrast, OffBa RF keeps high sensitivity and G-mean when the imbalance ratio increases. DyBa ORF achieves a sensitivity and specificity close to OffBa RF, but its update time is much less when new data arrives.

Interactive Segmentation of the Placenta and Adult Lungs. DyBa ORF was applied to two different 2D segmentation tasks: placenta segmentation of fetal MRI and adult lung segmentation of radiographs. Stacks of MRI images from 16 pregnancies in the second trimester were acquired with slice dimension 512×448, pixel spacing 0.7422mm×0.7422mm. A slice in the middle of each placenta was used, with the ground truth manually delineated by a radiologist.

Table 2. G-mean and Dice Score(DS) of DyBa ORF and counterparts in placenta and adult lung segmentation. G-mean and DS(RF) were measured on probability given by RFs. DS(CRF) was measured on the result after using CRF. t_u is the time for forests update after the arrival of new scribbles. The bold font shows values that are not significantly different from the corresponding results of OffBa RF(p-value>0.05).

	Method	G-mean(%)	DS(RF)(%)	DS(CRF)(%)	Average t_u(s)
Placenta	OffBa RF	84.24±4.02	74.97±7.20	89.32±3.62	1.80±0.92
	DyBa ORF	83.09±4.18	**75.25±6.88**	**89.17±3.73**	0.42±0.22
	MP ORF	78.21±8.12	71.98±9.76	85.14±9.13	0.37±0.18
	SP ORF	74.49±6.94	69.40±8.55	79.32±12.07	0.53±0.26
Adult Lungs	OffBa RF	90.80±2.30	86.87±3.89	94.25±1.62	7.40±1.17
	DyBa ORF	90.08±2.36	**86.69±3.56**	**94.06±1.64**	1.52±0.43
	MP ORF	85.51±3.82	82.95±4.43	90.53±3.59	1.14±0.30
	SP ORF	83.38±5.52	80.93±6.70	87.27±9.18	2.19±0.68

Lung images and ground truth[3] were downloaded from the JSRT Database[4]. Data from the first 20 normal patients were used (image size 2048×2048, pixel spacing 0.175mm×0.175mm) . At the start of segmentation, the user draws an initial set of scribbles to indicate the foreground and background and the RFs and CRF are applied. After that the user gives more scribbles several times and each time RFs are updated and used to predict the probability at each pixel.

Fig 2 shows an example of the placenta and adult lung segmentation with increasing scribbles. In Fig 2(a) and 2(b), lower accuracy of MP ORF and SP ORF compared with OffBa RF and DyBa ORF can be observed in the second and third column. Quantitative evaluations of both segmentation tasks after the last stage of interaction are listed in Table 2. We measured the G-mean and Dice score (DS) of the probability map thresholded by 0.5, DS after using CRF, and the average update time after the arrival of new scribbles. Table 2 shows DyBa ORF achieved a higher accuracy than MP ORF and SP ORF, and a comparable accuracy with OffBa RF, with largely reduced update time.

4   Discussion and Conclusion

Experiment results show that SP ORF had the worst performance, due to its lack of explicitly dealing with data imbalance. MP ORF [1] performed better, but it failed to be adaptive to imbalance ratio changes. OffBa RF, which learns from scratch for each update, and DyBa ORF, which considers the new imbalance ratio in both existing and new data, were adaptive to imbalance ratio changes. DyBa ORF's comparable accuracy and reduced update time compared with OffBa RF show that it is more suitable for interactive image segmentation. In addition, the results indicate that the MP/SP ORF needs some additional user interaction to achieve the same accuracy as obtained by DyBa ORF. This

---

[3] http://www.isi.uu.nl/Research/Databases/SCR/
[4] http://www.jsrt.or.jp/jsrt-db/eng.php

indirectly demonstrates that our model is helpful in reducing user interaction and saving interaction time. Future works can be done to further investigate the performance of DyBa ORF in reducing user interaction in segmentation tasks.

In conclusion, we present a dynamically balanced online random forest to deal with incremental and imbalanced training data with changing imbalance ratio, which occurs in the scribble-and-learning-based image segmentation. Our method is adaptive to imbalance ratio changes by combining a dynamically balanced online Bagging and a tree growing and shrinking strategy to update the random forests. Experiments show it achieved a higher accuracy than traditional ORF, with a higher efficiency than its offline counterpart. Thus, it is better for interactive image segmentation. It can also to be applied to other online learning problems with imbalanced data and changing imbalance ratio.

References

1. Barinova, O., Shapovalov, R., Sudakov, S., Velizhev, A.: Online Random Forest for Interactive Image Segmentation. In: EEML. pp. 1–8 (2012)
2. Boykov, Y., Jolly, M.P.: Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. ICCV 2001. 1(July), 105–112 (2001)
3. Breiman, L.: Random Forests. European Journal of Mathematics 45, 5–32 (2001)
4. Chen, C., Liaw, A., Breiman, L.: Using random forest to learn imbalanced data. University of California, Berkeley pp. 1–12 (2004)
5. Grady, L., Schiwietz, T., Aharon, S., Westermann, R.: Random walks for interactive organ segmentation in two and three dimensions: Implementation and validation. In: MICCAI 2005. vol. 3750, pp. 773–780 (2005)
6. Harrison, A.P., Birkbeck, N., Sofka, M.: IntellEditS: Intelligent learning-based editor of segmentations. In: MICCAI 2013. vol. 8151, pp. 235–242 (2013)
7. Saffari, A., Leistner, C., Santner, J., Godec, M., Bischof, H.: Online random forests. In: ICCV Workshops 2009. pp. 1393–1400 (2009)
8. Santner, J., Unger, M., Pock, T., Leistner, C., Saffari, A., Bischof, H.: Interactive Texture Segmentation using Random Forests and Total Variation. BMVC 2009 pp. 66.1–66.12 (2009)
9. Top, A., Hamarneh, G., Abugharbieh, R.: Active learning for interactive 3D image segmentation. In: MICCAI 2011. vol. 14, pp. 603–10 (2011)
10. Wang, G., Zuluaga, M.A., Pratt, R., Aertsen, M., David, A.L., Deprest, J., Vercauteren, T., Ourselin, S.: Slic-Seg: slice-by-slice segmentation propagation of the placenta in fetal MRI using one-plane scribbles and online learning. In: MICCAI 2015. pp. 29–37 (2015)
11. Yang, W., Cai, J., Zheng, J., Luo, J.: User-friendly interactive image segmentation through unified combinatorial user inputs. IEEE TIP 19(9), 2470–2479 (2010)