

LIGHT RANDOM REGRESSION FORESTS FOR AUTOMATIC MULTI-ORGAN LOCALIZATION IN CT IMAGES

Prasad N. Samarakoon*

Emmanuel Promayon*

Céline Fouard*

*Univ. Grenoble Alpes / CNRS / TIMC-IMAG UMR 5525, F-38041, France.

ABSTRACT

Classic Random Regression Forests (RRFs) used for multi-organ localization describe the random process of multivariate regression by storing the histograms of offset vectors along each bounding wall direction per leaf node. On the one hand, the RAM and storage requirements of classic RRFs may become exorbitantly high when such a RRF consists of many leaf nodes, but on the other hand, a large number of leaf nodes are required for better localization. We introduce Light Random Regression Forests (LRRFs) which eliminate the need to describe the random process by formulating the localization prediction based on the random variables that describe the random process. Consequently, LRRFs with the same localization capabilities require less RAM and storage space compared to classic RRFs. LRRF comprising 4 trees with 17 decision levels is approximately 9 times faster, takes 10 times less RAM, and uses 30 times less storage space compared to a similar classic RRF.

Index Terms— Random Regression Forests, Organ Localization

1. INTRODUCTION

In the field of medical image analysis, Multi-Organ Localization (MOL) is a preliminary step in both anatomical structure segmentation and registration. It also plays an important role in content specific image retrieval. Based on pioneering work of Breiman [1], Random Regression Forests (RRFs) were used in 2010 for the first time to solve the MOL problem by localizing the organs using bounding boxes [2]. The ingenuity of the proposal was the manner in which the MOL problem was transformed into a multivariate regression problem. It comprised the regression of a 6 Dimensional (6D) displacement vector (*offset vector*) from the bounding box walls of the organs to any given voxel. In 2013, the same authors proposed a modified implementation of RRFs [3, 4] that enhanced the previously achieved results by modifying the split node optimization method, the description of the random process, and the eventual usage of this description for prediction.

This work was supported by the French ANR within the TECS project Robacus (ANR-11-TECS-020-01).

Since its introduction, RRFs have been successfully used as the first localization step in fully automatic segmentation of right and left kidneys [5], liver [6], multiple organs (liver, right kidney, left kidney, spleen, gallbladder, and stomach) [7], hip joint [8], and detection, grading and classification of coronary stenoses [9] in CT volumes.

Predicting the localization of organs is based on using the *offset vectors* stored in the RRF leaf nodes during the training phase. The *offset vectors* $\mathbf{d}(\mathbf{v}; c)$ are computed for each organ and for each voxel using:

$$\mathbf{d}(\mathbf{v}; c) = \hat{\mathbf{v}} - \mathbf{b}(c), \quad (1)$$

where $c \in \text{organs}$ (the set of organs to localize), $\mathbf{b}(c)$ the bounding box vector, and $\hat{\mathbf{v}} = (v_x, v_x, v_y, v_y, v_z, v_z)$ made from the voxel position (v_x, v_y, v_z) . $\mathbf{d}(\mathbf{v}; c)$ is composed of 6 components (d_c^{dir}) where $\text{dir} \in \{\text{left, right, anterior, posterior, inferior, superior}\}$.

Given a set of voxels arriving at a particular node and for a given organ, the current setting of RRF regresses the continuous conditional distribution of $\mathbf{d}(\mathbf{v}; c)$ as a 6D multivariate Gaussian [2, 3, 4]. Consequently, the 6D multivariate Gaussian results in 6 1D univariate Gaussians.

The authors of [3, 4] described the random process attached to the continuous conditional distribution by storing the 1D histograms of *offset vectors* per each wall direction per each organ. Describing the random process results in a significant storage space requirement. In addition to that, these saved 1D histograms are loaded into the RAM during the prediction phase (also called the testing phase) in order to carry out the localization of an unseen image. This again translates into a substantial RAM requirement at the prediction phase.

In this article we present Light Random Regression Forests (LRRFs) which describe the random variables inherent to the random processes that were previously described in classic RRFs. By describing the random variables, the storage and RAM requirements are drastically reduced compared to the classic RRFs.

2. ON GAUSSIAN DISTRIBUTION SUMMATION

The foundation of LRRF is based on the following observation on summation of 1D Gaussian distributions. A 1D Gaussian distribution ($\mathcal{G}$) is described by its mean (μ) and variance

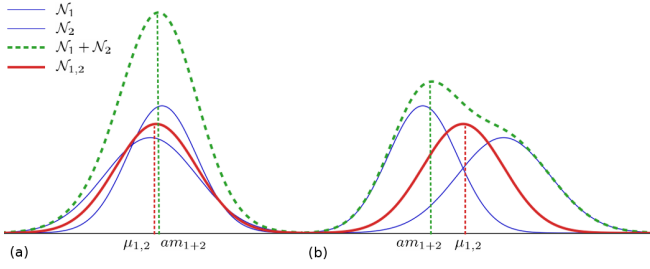


Fig. 1: $\mathcal{N}_1$ and $\mathcal{N}_2$ are two 1D Gaussian distributions (shown in blue). $\mathcal{N}_1 + \mathcal{N}_2$ is the distribution generated by adding $\mathcal{N}_1$ and $\mathcal{N}_2$ together (shown in green). $\mathcal{N}_{1,2}$ is another Gaussian distribution constructed from $\mu_{1,2}$: the weighted means (equal weights in this case) of $\mathcal{N}_1$ and $\mathcal{N}_2$ (shown in red). The goal is to check whether $\mu_{1,2}$ is a good approximation of the $\arg \max$ of $\mathcal{N}_1 + \mathcal{N}_2$ (am_{1+2}). (a) The $\arg \max$ can be closely estimated by the weighted means when the means of the original distributions are closer. (b) When the means of the original distributions are far apart, taking the $\arg \max$ privileges one distribution over the other.

(σ^2) in the following manner:

$$\mathcal{G} \sim \mathcal{N}(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (2)$$

A new distribution is made by summing many 1D Gaussian distributions together. The ultimate goal is to find the $\arg \max$ of the final distribution. This is similar to adding many prediction distributions described by histograms together and finding the $\arg \max$ of the final summed distribution in order to obtain the absolute bounding wall location in classic RRFs [3, 4].

Although the addition of two 1D Gaussians does not result in another 1D Gaussian, when the means of the distributions are very close, weighted means of the distributions are a good estimator of the $\arg \max$ of the summation (see Fig. 1a). But when the means of the distributions are quite apart from each other, the weighted means are not a good estimator of the $\arg \max$ (see Fig. 1b).

At the prediction phase of RRFs, “a good set” of leaf nodes that shows the lowest variability is selected per organ for the final localization prediction. Since each of these selected leaf node should predict approximately the correct position (hence, have approximately the same mean), we hypothesized that finding the $\arg \max$ of the final prediction summation distribution can be estimated using the weighted means of each selected leaf node.

3. LIGHT RANDOM REGRESSION FORESTS

Similarly to RRFs, LRRFs are also an ensemble of Random Regression Trees (RRTs). Each RRT consists of split and leaf nodes. Each split node divides the incoming voxels into two

separate child nodes. Leaf nodes are responsible for predicting the localization of organs.

The training phase of LRRFs is identical to the training phase of classic RRFs (see [3, 4] for a more detailed description) except for one simplification that occurs as a direct consequence of our hypothesis mention in Sect. 2. This simplification is described in Sect. 3.1 and the details of the prediction phase of LRRFs are presented in Sect. 3.2.

3.1. Training phase implementation

The split nodes are trained following the exact procedure of classic RRFs. When all the split nodes are trained, the leaf nodes contain all the accumulated voxels. Training of a given leaf node consists of summarizing the *offset vectors* of these accumulated voxels. As previously mentioned, RRFs regress the continuous conditional distribution of *offset vectors* as a 6D multivariate Gaussian which results in 6 1D univariate Gaussians along each bounding wall direction [2, 3, 4]. The classic RRFs train the leaf node by storing the random process using histograms of the *offset vectors* [3, 4].

Thanks to our hypothesis, we can eliminate the storing of the histograms (*i.e.* the description of the random process). Hence, per organ, we only store the mean (μ_{dir}) and the variance (σ_{dir}^2) of these 1D distributions (*i.e.* the random variables that define the distribution $\mathcal{D}$).

$$\mathcal{D} \sim \mathcal{N}(\mu_{dir}, \sigma_{dir}^2). \quad (3)$$

LRRFs are “light weight” by saving a lot of storage space as they do not save the 1D histograms of each wall direction of each organ.

3.2. Prediction phase implementation

During the prediction phase, a previously unseen image is put through an already trained LRRF in order to localize the organs. Similar to classic RRFs, each voxel ($\mathbf{v}$) of the image ($\mathcal{V}$) is pushed through each RRT until it is accumulated at a leaf node for each RRT of the LRRF ($l(\mathbf{v})$). The set of leaf nodes that accumulates at least 75% of voxels ($\tilde{\mathcal{L}}$) and that displays the lowest variability per each organ is used for the localization prediction similar to [3, 4]. Then, each bounding wall distribution ($\mathcal{D}(\mathbf{b}(c))$) can be defined as :

$$\mathcal{D}(\mathbf{b}(c)) \sim \sum_{l \in \tilde{\mathcal{L}}} \mathcal{D}(\mathbf{b}(c | l(\mathbf{v}))) \cdot \Pr(l), \quad (4)$$

where $\Pr(l) = |l(\mathbf{v})| / \sum_{l(\mathbf{v}) \in \tilde{\mathcal{L}}} |l(\mathbf{v})|$, $|l(\mathbf{v})|$ is the number of voxels accumulated at the leaf node l , and $\mathbf{b}(c | l(\mathbf{v}))$ can be derived from (1). From (3) this translates into an addition of weighted 1D Gaussians. The final prediction of the absolute position of each bounding wall direction ($\tilde{\mathbf{b}}_c$) is the $\arg \max$ of $\mathcal{D}(\mathbf{b}(c))$.

In classic RRFs, $\mathcal{D}(\mathbf{b}(c))$ is modeled by histograms per each bounding wall direction and $\tilde{\mathbf{b}}_c$ is obtained by finding the mode of the histogram (*i.e.* the $\arg \max$) [3, 4].

We use our hypothesis in order to find $\tilde{\mathbf{b}}_c$ by calculating the sum of the means of $\mathbf{b}(c | \mathbf{l}(\mathbf{v}))$ weighted by $\Pr(l)$ as:

$$\tilde{\mathbf{b}}_c = \sum_{l \in \tilde{\mathcal{L}}} \mu_b \cdot \Pr(l) , \quad (5)$$

where $\mu_b = \bar{\mathbf{v}}_{dir} - \mu_d$. LRRFs are “light” by leading to faster computational times as no random process is mimicked by the use of histograms but a model of $\tilde{\mathbf{b}}_c$ is directly built using the random variables.

4. MATERIALS AND EXPERIMENTS

Our dataset comprised 100 anonymized CT volumes belonging to 100 different patients that were obtained from the teaching hospital of Grenoble Alpes. These CT volumes showed a high variability in size (from $263 \times 263 \times 366$ to $466 \times 466 \times 568$ in mm^3), imaged region, noise level, with or without the injection of contrast agents, presence or absence of biomedical implants, *etc.* The dataset was randomly divided, similarly to [2], into a training set and a testing set (or prediction set) consisting of 55 and 45 volumes respectively.

The left kidney, right kidney, liver, spleen, left femur head, right femur head, left pelvis, right pelvis, and L5 vertebra were used for this study. All organs of interest of all 100 images were manually delineated using 3D bounding boxes. These manually delineated bounding boxes were used as the gold standard.

A classic RRF was trained and used for prediction as described in [3, 4]. Then, the classic RRF was transformed into a LRRF by removing the stored 1D histograms of offset vectors per each wall direction as mentioned in Sect. 3. This enabled us to compare the two methods under the exact same conditions.

We generated two results sets from classic RRFs (R_{C1}) and from LRRFs (R_{L1}); namely:

- R_{C1} and R_{L1} with 4 RRTs and $D = 12$ similar to [3, 4],
- R_{C2} and R_{L2} with 4 RRTs and $D \in [1, 2, \dots, 17]$,

to evaluate the precision and usability of both methods.

Validation of our hypothesis: In order to evaluate the hypothesis of estimating the $\arg \max$ of summation of 1D Gaussians by the weighted summation of their means as presented in Sect. 2, the following steps were carried out. First, the distribution pairs of $\arg \max$ of $p(\mathbf{b}_c)$ and weighted means of $p(\mathbf{b}_c | l)$ denoted respectively by $\mathcal{D}_a$ and $\mathcal{D}_\mu$ were created for each wall direction per organ (6×9 distribution sets) using all 45 prediction images. Finally, each pair of $\mathcal{D}_a$ and $\mathcal{D}_\mu$ were subjected to Wilcoxon–Mann–Whitney test.

Prediction precision: Bounding Wall Prediction Error (BWPE) was used as the performance evaluation metric similar to [2, 3]. BWPE is the absolute difference between the

predicted and gold standard bounding box walls. As there are six walls per bounding box, the mean BWPE (Er) is calculated taking the mean of the six BWPEs. Er was calculated only for the organs that were fully present in the CT volumes.

Usability: The usability was evaluated using the mean time (t) and the mean amount of RAM (M) required to do a MOL in a CT volume. The amount of disk space used to store the forest (S) was also compared.

For all statistical tests carried out, α was set to 0.01. All results were generated using a C++ implementation on a quad-core Intel® Xeon® 3GHz machine with 32GB of RAM.

5. RESULTS

Hypothesis verification: Wilcoxon–Mann–Whitney test failed to reject the null hypothesis of $\mathcal{D}_a$ and $\mathcal{D}_\mu$ coming from the same underlying distribution with p -value $\in [0.04 - 0.98]$ for all 54 verification distribution pairs of $\mathcal{D}_a$ and $\mathcal{D}_\mu$. This provides strong empirical evidence that the $\arg \max$ of summation of 1D Gaussians can be readily estimated by the weighted summation of their means.

Prediction precision evaluation: The mean BWPEs (Er_1) obtained for R_{C1} and R_{L1} per each organ are presented in Fig. 2a. The obtained Er_1 failed to reject the null hypothesis that both error measures of R_{C1} and R_{L1} for all organs originated from the same underlying distribution with p -value $\in [0.03 - 0.78]$. That indicates the error is statistically the same with the two methods.

The mean BWPEs (Er_2) of all organs obtained were in the range of $[9.85 - 16.53]$ for R_{C2} and in the range of $[9.99 - 16.23]$ for R_{L2} . The standard deviations of R_{C2} and R_{L2} fell between $[10.21 - 15.17]$ and $[9.59 - 14.23]$ respectively. Er_2 decreased with the number of decision levels for both R_{C2} and R_{L2} .

Usability evaluation: The metrics t_2 , M_2 , and S_2 for R_{C2} and R_{L2} are presented in Fig. 2b. Both t_2 and S_2 for R_{C2} grow exponentially compared to the same values of R_{L2} . For 17 decision levels, t_2 , M_2 , and S_2 are (2.2s, 19.5s), (117 MB, 1147 MB), and (171 MB, 5221 MB) for R_{L2} and R_{C2} respectively. MOL using the LRRF is approximately 9 times faster, takes about 10 times less RAM, and uses about 30 times less storage space compared to the classic RRF.

6. DISCUSSION AND CONCLUSION

Although estimating the $\arg \max$ of summation of 1D Gaussians by weighted individual means might be questionable (as the sum of 1D Gaussians is not a Gaussian distribution), the empirical results indicated that they are statistically alike.

From R_{C2} and R_{L2} , it is apparent that as the number of decision levels increase, the feasibility of classic RRFs for MOL becomes questionable. We propose LRRFs as an alternative that only describes the random variables that are inher-

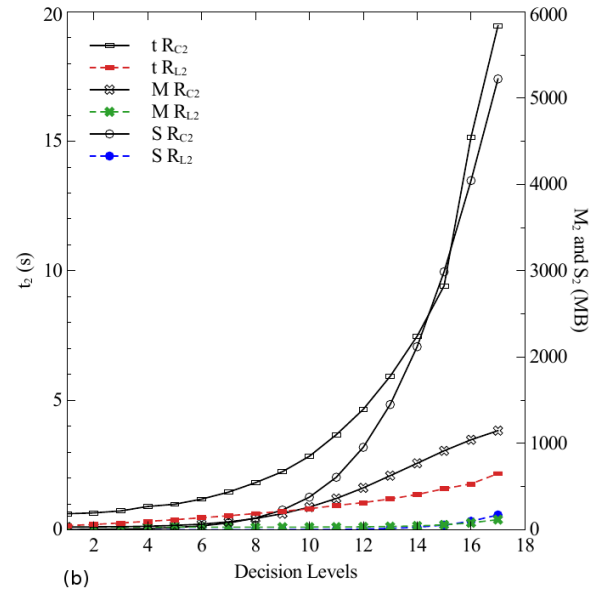
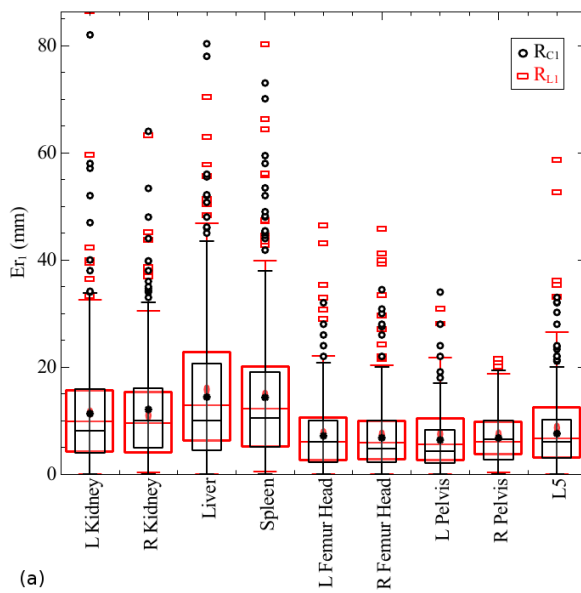


Fig. 2: Prediction precision and usability evaluation. (a) 1.5 Inter quartile range box plots of mean BWPEs per organ for R_{C1} and R_{L1} . (b) Evolution of t_2 , M_2 , and S_2 with number of decision levels for R_{C2} and R_{L2} .

ent to the random process which results in a huge RAM and storage requirement reduction enabling the growth of random forests having deeper trees.

The huge gains in speed and memory obtained using LRRFs without compromising the localization capabilities enable the use of RRFs in resource limited environments such as mobile phones or smart wearables. Another advantage of LRRF is the ability to transform a classic RRF into a LRRF and use it in a resource critical environment.

7. REFERENCES

- [1] Leo Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [2] Antonio Criminisi, Jamie Shotton, Duncan Robertson, and Ender Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies," in *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pp. 106–117. Springer, 2010.
- [3] Antonio Criminisi, Duncan Robertson, Ender Konukoglu, Jamie Shotton, Sayan Pathak, Steve White, and Khan Siddiqui, "Regression forests for efficient anatomy detection and localization in computed tomography scans," *Medical image analysis*, vol. 17, no. 8, pp. 1293–1303, 2013.
- [4] Antonio Criminisi, D Robertson, O Pauly, B Glocker, E. Konukoglu, Jamie Shotton, D Mateus, A Martinez Möller, S G Nekolla, and N Navab, "Anatomy detection and localization in 3D medical images," in *Decision forests for computer vision and medical image analysis*, Antonio Criminisi and Jamie Shotton, Eds., chapter 14, pp. 193–209. Springer, 2013.
- [5] Rémi Cuingnet, Raphael Prevost, David Lesage, Laurent D Cohen, Benoît Mory, and Roberto Ardon, "Automatic detection and segmentation of kidneys in 3D CT images using random forests," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012*, pp. 66–74. Springer, 2012.
- [6] Romane Gauriau, Rémi Cuingnet, Raphael Prevost, Benoit Mory, Roberto Ardon, David Lesage, and Isabelle Bloch, "A generic, robust and fully-automatic workflow for 3D CT liver segmentation," in *Abdominal Imaging. Computation and Clinical Applications*, pp. 241–250. Springer, 2013.
- [7] Romane Gauriau, Rémi Cuingnet, David Lesage, and Isabelle Bloch, "Multi-organ localization combining global-to-local regression and confidence maps," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*, pp. 337–344. Springer, 2014.
- [8] Chengwen Chu, Cheng Chen, Li Liu, and Guoyan Zheng, "Facts: fully automatic ct segmentation of a hip joint," *Annals of biomedical engineering*, vol. 43, no. 5, pp. 1247–1259, 2015.
- [9] B Michael Kelm, Sushil Mittal, Yefeng Zheng, Alexey Tsymbal, Dominik Bernhardt, Fernando Vega-Higuera, S Kevin Zhou, Peter Meer, and Dorin Comaniciu, "Detection, grading and classification of coronary stenoses in computed tomography angiography," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011*, pp. 25–32. Springer, 2011.