Regression Forests for Efficient Anatomy Detection and Localization in Computed Tomography Scans

A. Criminisi[†], D. Robertson[†], E. Konukoglu[†], J. Shotton[†], S. Pathak[‡], S. White[‡], and K. Siddiqui[‡]

[†]*Microsoft Research Ltd, Cambridge, UK.*
[‡]*Microsoft Corporation, Redmond, WA, USA.*

Abstract

This paper proposes a new algorithm for the efficient, automatic detection and localization of multiple anatomical structures within three-dimensional computed tomography (CT) scans. Applications include selective retrieval of patients images from PACS systems, semantic visual navigation and tracking radiation dose over time.

The main contribution of this work is a new, continuous parametrization of the anatomy localization problem, which allows it to be addressed effectively by *multi-class random regression forests*. Regression forests are similar to the more popular classification forests, but trained to predict *continuous*, multi-variate outputs, where the training focuses on maximizing the confidence of output predictions. A single pass of our probabilistic algorithm enables the direct mapping from voxels to organ location and size.

Quantitative validation is performed on a database of 400 highly variable CT scans. We show that the proposed method is more accurate and robust than techniques based on efficient multi-atlas registration and template-based nearest-neighbour detection. Due to the simplicity of the regressor's context-rich visual features and the algorithm's parallelism, these results are achieved in typical run-times of only $\sim$4 seconds on a conventional single-core machine.

1. Introduction

This paper proposes a new, parallel algorithm for the efficient detection and localization of anatomical structures ('organs') in 3D computed tomography studies. Localizing anatomical structures is an important step for

many subsequent image analysis tasks (possibly organ-specific) such as segmentation, registration and classification. It is also crucial for managing database systems and creating intelligent navigation and visualization tools. For instance, one application is the efficient retrieval of selected portions of patients' scans from PACS databases. When a physician wishes to inspect a particular organ, the ability to determine its position and extent automatically means that it is not necessary to retrieve the entire scan (which could comprise hundreds of MB of data) but a smaller region of interest. Thus it is possible to achieve faster user interaction while making economical use of the limited bandwidth. The proposed organ localizer could potentially be used also for tracking the amount of radiaton absorbed by each organ over time. However, in its current form, the approximate representation of organs would produce indicative dose estimations.

The main contribution of this work is a new parametrization of the anatomy localization task as a multivariate, continuous parameter estimation problem. This is addressed effectively via tree-based, non-linear regression. Unlike the popular *classification* forests (often referred to simply as "random forests"), *regression* forests (Breiman et al., 1984) have not yet been used in medical image analysis. Our approach is fully probabilistic and, unlike previous techniques, *e.g.* (Zhou et al., 2007; Fenchel et al., 2008), is trained to maximize the confidence of output predictions. As a by-product, our method produces *salient anatomical landmarks*; *i.e.* automatically selected "anchor" regions that help localize organs of interest with high confidence. Our algorithm can localize both macroscopic anatomical regions[1] (*e.g.* abdomen, thorax, trunk, *etc.*) and smaller scale structures (*e.g.* heart, l. adrenal gland, femoral neck, *etc.*) using a single, efficient model, *c.f.* (Feulner et al., 2009).

Motivated mostly by the semantic navigation use-case scenario, our focus in this paper is on both accuracy of prediction and speed of execution. Our goal is to achieve accurate anatomy localization in seconds on a conventional machine.

1.1. Literature review

Regression approaches. Regression algorithms (Hardle, 1990) estimate functions which map input variables to *continuous* outputs[2]. The regression

[1] This is useful because the existing anatomical region DICOM tag is often inaccurate (Gueld et al., 2002).

[2] as opposed to *classification* where the predicted variables are discrete, categorical.

paradigm fits the anatomy localization task well. In fact, its goal is to learn the non-linear mapping from voxels *directly* to organ position and size.

The first work to use regression for anatomy localization in images is Zhou et al. (2005). There, the authors need to define the non-linear mapping as an analytical function whose exact form is learned via regularized boosting. They also present a thorough overview of different regression techniques and discuss the superiority of boosted regression. In their later work (Zhou et al., 2007), their boosted regression technique was improved by incorporating high degree-of-freedom weak learners. The main difference between that approach and the one presented here is in the non-linear mapping. Defining a regression function analytically as done in Zhou et al. (2005, 2007) has two major drawbacks: 1) the definition of the function requires critical modeling assumptions for the type of the weak learner and the regularization term, and 2) obtaining a confidence measure for the regression output is non trivial. In contrast, our approach does not assume an analytical form for the mapping. This results in a simpler formulation with fewer modeling choices. In addition, the probabilistic nature of our method yields a natural way of associating confidence with the predicted output. In fact, the training phase of our algorithm directly maximizes the confidence of the predicted probability distribution.

A comparison between boosting, forests and cascades is found in Yin et al. (2007). To our knowledge, so far only two papers have used regression forests in imaging (Montillo and Ling, 2009; Gall and Lempitsky, 2009), neither with application to medical image analysis. For instance, Gall and Lempitsky (2009) address the problem of detecting pedestrians *vs.* background. For the readers who might not be familiar with regression forests we provide a short explanation in the appendix. Also, a detailed description of general decision forests and their applications may be found in Criminisi and Shotton (2013), with free research code and demos available at `http://research.microsoft.com/projects/decisionforests`.

Classification-based approaches. In Zhan et al. (2008) organ detection is achieved via a confidence maximizing sequential scheduling of multiple, organ-specific *classifiers*. In contrast, our single, tree-based regressor allows us to deal naturally with multiple anatomical structures simultaneously. As shown in the machine learning literature (Torralba et al., 2007) this encourages feature sharing and, in turn better generalization. In Seifert et al. (2009) a sequence of probabilistic boosting tree (PBT) classifiers (first for salient slices,

then for landmarks) are used. In constrast, our single regressor maps directly from voxels to organ poses; latent, salient landmark regions are extracted as a by-product. In Criminisi et al. (2009) the authors achieve localization of organ *centres* but fail to estimate the organ extent (similar to Gall and Lempitsky (2009)). Here we present a more direct, continuous model which estimates the position of the walls of the bounding box containing each organ; thus achieving simultaneous organ localization and extent estimation.

Marginal Space Learning. One of the most popular approaches for object localization in medical images is Marginal Space Learning (MSL) proposed in Zheng et al. (2007, 2009a). MSL has been demonstrated to be very useful in practice (Zheng et al., 2009b; Barbu et al., 2012). However, that algorithm has three limitations. Firstly, MSL is designed to detect a *single* object at a time and extending it to the joint-localization of multiple objects (*e.g.* more than 20) is not immediate. For example, existing extensions rely on applying the algorithm iteratively, one run for each object of interest. The order of detection is either determined through combinatorial optimization or driven by the confidence values each object attains during the detection phase (Liu et al., 2010). In contrast, our method achieves joint-localization of any number of structures without modification and without worrying about complex ordering strategies.

Secondly, MSL builds upon multiple classification stages. For instance, to detect the position of the heart we may need: 1) a classifier trained to estimate overall translation, 2) a classifier trained on translation and rotation, and 3) yet another classifier trained on translation, rotation and scale. All three classifiers need be applied for each organ in a sequence. For *e.g.* 20 organs we would need to train $20 \times 3 = 60$ different classifiers, with clear scalability issues. In contrast, we propose using a single forest regressor (with *e.g.* only $\sim$4 trees) to deal with multiple organs (here tested on 26 anatomical structures).

Thirdly, we argue that solving a localization problem via classification is not optimal. In MSL, *binary* classifiers are run in a sliding-window fashion. For each point the classifier produces a positive answer (point is "close" to the structure) or a negative one (point is "far" from the structure). But reducing real-valued distances to binary decisions introduces a loss. Also, defining positive and negative examples is an ambiguous task. Instead, our regression forest directly estimates the 3D displacement of each voxel from the target regions. On the flip side, it is also true that in practice learning

good classifiers seems to be easier than learning good regressors. This may be due to the fact that as a community we have had much more exposure to classification tasks than regression ones. This paper shows that for the application of anatomical bounding box localization using a regression forest can be more accurate than using a classification approach.

Registration-based approaches. Although atlas-based methods have enjoyed much popularity (Fenchel et al., 2008; Shimizu et al., 2006; Yao et al., 2006), their conceptual simplicity belies the technical difficulty inherent in achieving robust, inter-subject registration. Robustness may be improved by using multi-atlas techniques (Isgum et al., 2009) but only at the expense of multiple registrations and hence increased computation time. Our algorithm incorporates atlas information within a compact tree-based model. As shown in the results section, such model is more efficient than keeping around multiple atlases and achieves anatomy localization in only a few seconds. Comparisons with global affine atlas registration methods (similar to ours in computational cost) show that our algorithm produces lower errors and more stable predictions. Next we describe details of our approach.

2. Multivariate regression forests for organ localization

This section presents mathematical notation, problem parametrization and other details of our multi-organ regression forest with application to anatomy localization in CT images.

Mathematical notation. Vectors are represented in boldface (*e.g.* $\mathbf{v}$), matrices as teletype capitals (*e.g.* $\mathtt{A}$), and sets in calligraphic style (*e.g.* $\mathcal{S}$). The position of a voxel in a CT volume is denoted $\mathbf{v} = (v_x, v_y, v_z)$.

The labelled database. The 26 anatomical structures we wish to recognize are $\mathcal{C} = \{$ `abdomen`, `l. adrenal gland`, `r. adrenal gland`, `l. clavicle`, `r. clavicle`, `l. femoral neck`, `r. femoral neck`, `gall bladder`, `head of l. femur`, `head of r. femur`, `heart`, `l. atrium of heart`, `r. atrium of heart`, `l. ventricle of heart`, `r. ventricle of heart`, `l. kidney`, `r. kidney`, `liver`, `l. lung`, `r. lung`, `l. scapula`, `r. scapula`, `spleen`, `stomach`, `thorax`, `thyroid gland`$\}$. We are given a database of 400 scans which have been manually annotated with 3D bounding boxes tightly drawn around the structures of interest (see fig. 2a). The bounding box for the organ $c \in \mathcal{C}$ is parametrized as a 6-vector $\mathbf{b}_c = (b_c^{\mathtt{L}}, b_c^{\mathtt{R}}, b_c^{\mathtt{A}}, b_c^{\mathtt{P}}, b_c^{\mathtt{H}}, b_c^{\mathtt{F}})$ where each element represents the position

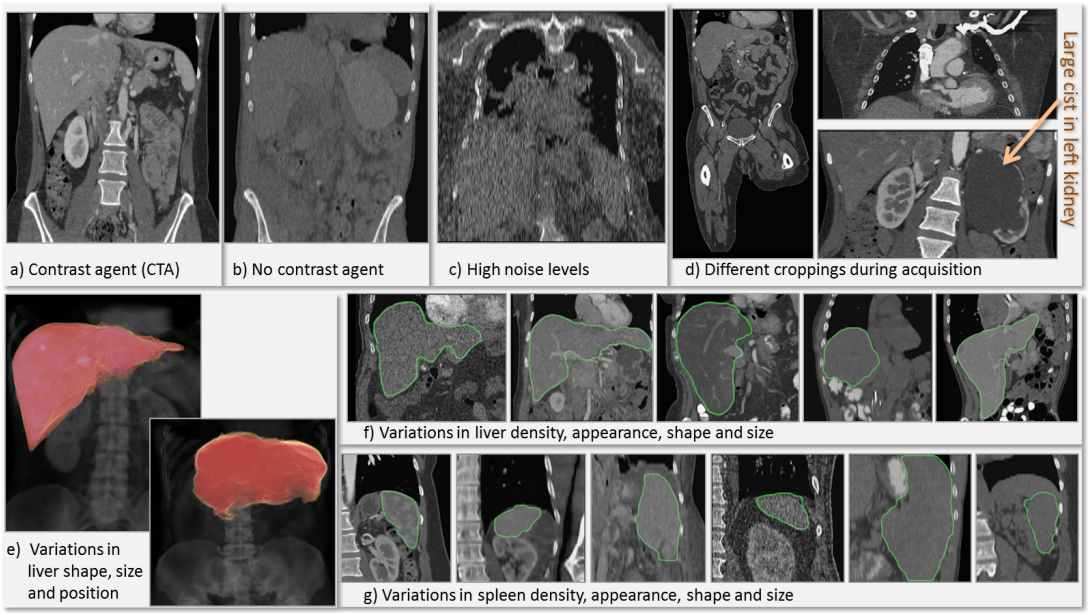


Figure 1: **Variability in our labelled database. (a, b, c)** Variability in appearance due to presence of contrast agent, or noise. **(d)** Difference in image geometry due to acquisition parameters and possible anomalies. **(e)** Volumetric renderings of liver and spine to illustrate large changes in their relative position and in the liver shape. **(f,g)** Mid-coronal views of liver and spleen across different scans in our database to illustrate their variability. All views are metrically and photometrically calibrated.

(in mm) of one axis-aligned face[3]. The database comprises patients with a wide variety of medical conditions and body shapes and the scans exhibit large differences in image cropping, resolution, scanner type, and use of contrast agents (fig. 1). Voxel sizes are $\sim 0.5 - 1.0$ mm along x and y, and $\sim 1.0 - 5.0$ mm along z. The images have not been pre-registered or normalized in any way. The goal is to localize organs of interest accurately and automatically, despite such large variability. The following sections describe how this is achieved.

2.1. Problem parametrization and regression forest learning

Key to our algorithm is the idea that *all* voxels in a test CT volume contribute with varying confidence to estimating the position of *all* organs'

[3]Superscripts follow standard radiological orientation convention: `L` = left, `R` = right, `A` = anterior, `P` = posterior, `H` = head, `F` = foot.

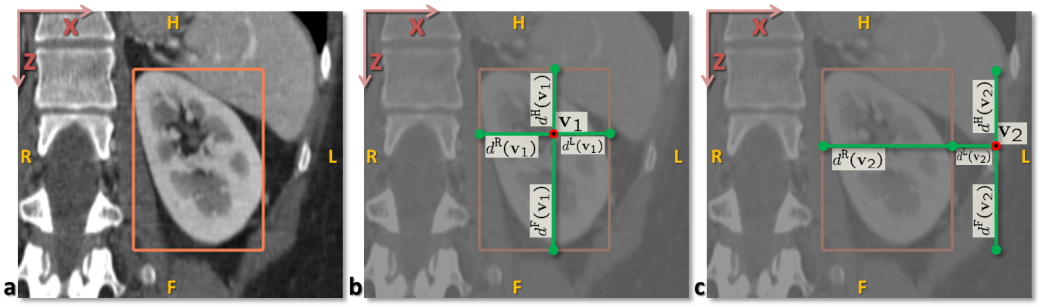


Figure 2: **Problem parametrization. (a)** A coronal view of a left kidney and the associated ground-truth bounding box (in orange). **(b, c)** *Every* voxel $\mathbf{v}_i$ in the volume votes for the position of the six walls of each organ's 3D bounding box via 6 relative, offset displacements $d^{\mathtt{k}}(\mathbf{v}_i)$ in the three canonical directions x, y and z.

bounding boxes (see fig. 2b,c). Intuitively, some distinct voxel clusters (*e.g.* ribs or vertebrae) may predict the position of an organ (*e.g.* the heart) with high confidence. Thus, at detection time those clusters should be used as landmarks for the localization of those organs. Our aim is to learn to cluster voxels based on their appearance, their spatial context and, above all, their confidence in predicting the position and size of all organs of interest. We tackle this simultaneous feature selection and parameter regression task with a multi-class random regression forest (fig. 3); *i.e.* an ensemble of regression trees trained to predict the location and size of all desired organs simultaneously. The desired output is one six-dimensional vector $\mathbf{b}_c$ per organ, a total of $6|\mathcal{C}|$ continuous parameters.

Note that this is very different from the task of assigning a categorical label to each voxel (*i.e.* the classification approach in Criminisi et al. (2009)). Here we wish to produce confident predictions of a small number of continuous localization parameters. The *latent* voxel clusters are discovered automatically without ground-truth cluster labels.

2.1.1. Forest training

The training process constructs each regression tree and decides at each node how to best split the incoming voxels. We are given a subset of all labelled CT volumes (the training set), and the associated ground-truth organ bounding box positions (fig. 2a). A subset of voxels in the training volumes is used for forest training. These training voxels are sampled on a regular grid within ±10 cm of the centre of each axial slice in the training volume.

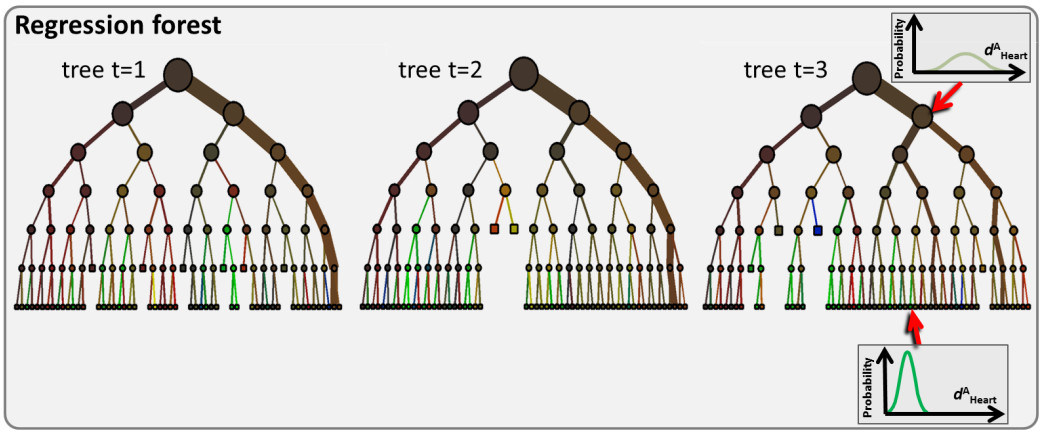


Figure 3: **A regression forest** is an ensemble of different regression trees. Each leaf contains a distribution for the continuous output variable/s. Leaves have associated different degrees of confidence (illustrated by the "peakiness" of distributions).

The size of the forest T is fixed and all trees are trained in parallel.

Each training voxel is pushed through each of the trees starting at the root. Each split node applies the following binary test $\xi_j > f(\mathbf{v};\boldsymbol{\theta}_j) > \tau_j$ and based on the result sends the voxel to the left or right child node. $f(.)$ denotes the feature response computed for the voxel $\mathbf{v}$. The parameters $\boldsymbol{\theta}_j$ describe the visual feature that is computed at the j^{th} node. Our visual features are similar to those in Gall and Lempitsky (2009); Criminisi et al. (2009); Shotton et al. (2009), *i.e.* mean intensities over displaced, asymmetric cuboidal regions of the volume. These features are efficient to compute and capture spatial context. The feature response is $f(\mathbf{v};\boldsymbol{\theta}_j) = |F_1|^{-1}\sum_{\mathbf{q}\in F_1} I(\mathbf{q}) - |F_2|^{-1}\sum_{\mathbf{q}\in F_2} I(\mathbf{q})$; with F_i indicating 3D box regions and I the intensity. F_2 can be the empty set for unary features. Randomness is injected at training time by making available at each node only a random sample of all possible features. This technique has been shown to increase the generalization of tree-based predictors (Ho, 1998). Next we discuss how to select the splitting function associated with each internal node.

Node optimization. Each voxel $\mathbf{v}$ in each training volume is associated with an offset with respect to the bounding box $\mathbf{b}_c$ for each class $c \in \mathcal{C}$ (see fig. 2b,c). Such offset is a function of both $\mathbf{v}$ and c as follows $\mathbf{d}(\mathbf{v};c) = \hat{\mathbf{v}} - \mathbf{b}_c(\mathbf{v})$, with $\hat{\mathbf{v}} = (v_x, v_x, v_y, v_y, v_z, v_z)$. Therefore $\mathbf{d}(\mathbf{v};c) \in \mathbb{R}^6$.

As with the training of classification trees, node optimization is driven by

maximizing an information gain measure, defined in general terms as: $H(\mathcal{S}) - \sum_{i=\{\mathtt{L},\mathtt{R}\}} \omega_i H(\mathcal{S}_i)$ where H denotes entropy, $\mathcal{S}$ is the set of training points reaching a node and $\mathtt{L}, \mathtt{R}$ denote its left and right children. In classification problems the entropy is defined over distributions of discrete class labels. In the context of regression, however, we measure the purity of the probability density of the real-valued predictions instead.

For a given class c we model the continuous conditional distribution of the vector $\mathbf{d}(\mathbf{v}; c)$ at each node as a multivariate Gaussian; *i.e.*

$$p(\mathbf{d}|c; \mathcal{S}) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathtt{\Lambda}_c(\mathcal{S})|} e^{-\frac{1}{2}(\mathbf{d}-\overline{\mathbf{d}}_c)^\top \mathtt{\Lambda}_c(\mathcal{S})^{-1}(\mathbf{d}-\overline{\mathbf{d}}_c)},$$

with $N = 6$ and $\int_{\mathbb{R}^6} p(\mathbf{d}|c; \mathcal{S})\ d\mathbf{d} = 1$. The vector $\overline{\mathbf{d}}_c$ indicates the mean displacement and the matrix $\mathtt{\Lambda}_c$ the covariance of $\mathbf{d}$ for all points in $\mathcal{S}$.

For the set $\mathcal{S}$ we also know the discrete class prior $p(c; \mathcal{S}) = n_c(\mathcal{S})/Z$, where $n_c(\mathcal{S})$ is the number of training voxels in the set $\mathcal{S}$ for which it is possible to compute the displacement $\mathbf{d}(\mathbf{v}; c)$; *i.e.* the training points in the set that come from training volumes for which the organ c is present. Z is a normalization constant such that $\sum_c p(c; \mathcal{S}) = 1$.

Thus we know the joint distribution $p(\mathbf{d}, c; \mathcal{S}) = p(\mathbf{d}|c; \mathcal{S})p(c; \mathcal{S})$. For a generic Gaussian-distributed random variable $\mathbf{x} \in \mathbb{R}^N$ with covariance $\mathtt{\Lambda}$ the differential entropy can be shown to be $H(\mathbf{x}) = \frac{1}{2} \log\left((2\pi e)^N |\mathtt{\Lambda}|\right)$. This leads (after algebraic manipulation) to the following joint entropy for the node:

$$H(\mathbf{d}, c; \mathcal{S}) = H(c; \mathcal{S}) + \sum_c p(c; \mathcal{S}) \left(\frac{1}{2} \log\left((2\pi e)^N |\mathtt{\Lambda}_c(\mathcal{S})|\right) \right) \tag{1}$$

The joint information gain is $IG = H(\mathbf{d}, c; \mathcal{S}) - \sum_{i \in \{L,R\}} \omega_i H(\mathbf{d}, c; \mathcal{S}_i)$ which after some manipulation can be rewritten as

$$IG = IG_d + IG_c \tag{2}$$

where

$$IG_d = \frac{1}{2} \left(\sum_c p(c; \mathcal{S}) \log |\mathtt{\Lambda}_c(\mathcal{S})| - \sum_{i \in \{L,R\}} \omega_i \sum_c p(c; \mathcal{S}_i) \log |\mathtt{\Lambda}_c(\mathcal{S}_i)| \right) \tag{3}$$

and

$$IG_c = H(c; \mathcal{S}) - \sum_{i \in \{L,R\}} \omega_i H(c; \mathcal{S}_i) \tag{4}$$

with $\omega_i = |\mathcal{S}_i|/|\mathcal{S}|$ the ratio of the number of points reaching the i^{th} child.

Maximizing (2) implies minimizing the determinants of the 6×6 covariance matrices associated with the $|\mathcal{C}|$ organs; where each organ's contribution is weighted by the associated prior probability. This decreases the uncertainty in the probabilistic vote cast by each cluster of voxels on each organ pose. In our experiments we found that this prior-driven organ weighting produces more balanced trees and has a noticeable effect on the accuracy of the results.

Branching stops when the number of points reaching the node is fewer than a threshold n_{min} or a maximum tree depth D has been reached (here $n_{min} = 25$ and $D = 12$). After training, the j^{th} decision node remains associated with the feature $\boldsymbol{\theta}_j$ and thresholds ξ_j, τ_j. At each leaf node we store the learned means $\overline{\mathbf{d}}_c$ and covariance matrices Λ_c, and the class priors $p(c)$, (fig. 3b).

This framework may be reformulated using non-parametric distributions, with pros and cons in terms of regularization and storage. We have found our parametric assumption not to be restrictive since the multi-modality of the input space is captured by our hierarchical piece-wise Gaussian model. However, under the simplifying assumption that bounding box face positions are uncorrelated (*i.e.* diagonal Λ_c), it is convenient to store at each leaf node learned 1D histograms over face offsets $p(\mathbf{d}|c; \mathcal{S})$.

Discussion. Equation (2) is an information-theoretical way of maximizing the confidence of the desired continuous output *for all* organs, without going through intermediate voxel classification (as in Criminisi et al. (2009) where difficult to define positive and negative examples of organ centres are needed). Furthermore, this gain formulation enables testing different context models; *e.g.* imposing a *full* covariance Λ_c would allow correlations between all walls in each organs. One could also think of enabling correlations between different organs. Taken to the extreme, this could have undesired over-fitting consequences. On the other hand, assuming *diagonal* Λ_c matrices leads to uncorrelated output predictions. Interesting models live in the middle ground, where some but not all correlations are enabled to capture *e.g.* class hierarchies or other forms of spatial context. For a more detailed description of forests training and associated code see Criminisi and Shotton (2013) and `http://research.microsoft.com/projects/decisionforests`.

2.1.2. Forest testing

Given a previously unseen CT volume $\mathcal{V}$, test voxels are sampled in the same manner as at training time. Each test voxel $\mathbf{v} \in \mathcal{V}$ is pushed through each tree starting at the root and the corresponding sequence of tests applied. The voxel stops when it reaches its leaf node $l(\mathbf{v})$, with l indexing leaves across the whole forest. The stored distribution $p(\mathbf{d}_c|l)$ for class c also defines the posterior for the absolute bounding box position: $p(\mathbf{b}_c|l)$ since $\overline{\mathbf{b}}_c(\mathbf{v}) = \hat{\mathbf{v}} - \overline{\mathbf{d}}_c(\mathbf{v})$. The posterior probability for $\mathbf{b}_c$ is now given by

$$p(\mathbf{b}_c) = \sum_{t=0}^{T} \sum_{l \in \tilde{\mathcal{L}}_t} p(\mathbf{b}_c|l) p(l). \tag{5}$$

$\tilde{\mathcal{L}}_t$ is a subset of the leaves of tree t. We select $\tilde{\mathcal{L}}_t$ as the set of leaves which have the smallest uncertainty (for each class c) and contain 75% of all test voxels. Finally $p(l)$ is simply the proportion of samples arriving at leaf l.

Organ localization. The final prediction $\tilde{\mathbf{b}}_c$ for the absolute position of the c^{th} organ is given by:

$$\tilde{\mathbf{b}}_c = \arg\max_{\mathbf{b}_c} p(\mathbf{b}_c). \tag{6}$$

Under the assumption of uncorrelated output predictions for bounding box faces, it is convenient to represent the posterior probability $p(\mathbf{b}_c)$ as six 1D histograms, one per face. We aggregate evidence into these histograms from the leaf distributions $p(\mathbf{b}_c|l)$. Then $\tilde{\mathbf{b}}_c$ is determined by finding the histogram maxima. Furthermore, we can derive a measure of the confidence of this prediction by fitting a 6D Gaussian with diagonal covariance matrix $\tilde{\mathbf{\Lambda}}$ to the histograms in the vicinity of $\tilde{\mathbf{b}}_c$. A useful measure of the confidence of the prediction is then given by $|\tilde{\mathbf{\Lambda}}|^{-1/2}$.

Organ detection. The organ c is declared present in the scan if the prediction confidence is greater than β. The parameter β is tuned to achieve the desired trade-off between the relative proportions of false positive and the false negative detections, and it is application dependent.

3. Results, comparisons and validation

This section assesses the proposed algorithm in terms of accuracy, runtime speed, and memory efficiency; and compares it to alternative techniques.

3.1. A simple 2D application

We begin with a 2D example which enables us to visualize intermediate results and helps our understanding of the algorithm. We have a database of 186 coronal images from different CT scans. The images are metrically and photometrically calibrated and for each the ground truth bounding box for the right kidney has been manually marked. The database has been split randomly into 50% training and 50% testing and a regression forest trained on the training set. Fig. 4 shows the results of applying the learned forest to previously unseen test images. The estimated kidney bounding box (in red) is close to the ground truth one (in blue). The forest also associates uncertainty (shown as a shaded red band) with the estimated output. Visualizing the learned trees (fig. 4a) confirms that as one goes down from the root towards the leaves the associated confidence for the predicted location of the walls increases. Finally, this 2D example enables us to visualize the computed landmark regions. In the figure we plot the test pixels which end up in the leaves with smallest variance for each box wall. As hoped we find that distinct anatomical regions such as the top of the lungs or the spine are automatically selected to predict the vertical and horizontal position of the kidney, respectively. A mix of many landmark locations are used with the relative weights automatically estimated. The presence of a large anomaly such as a collapsed lung does affect the accuracy of the localization though the final results is still acceptable (fig. 4f). Box wall localization errors are:

	L	R	H	F
error	5.64mm	5.96mm	3.84mm	6.88mm

Next we assess our actual organ localization algorithm on 3D CT images.

3.2. Accuracy in 3D anatomy localization

A regression forest was trained using 318 CT volumes selected randomly from our 400-volume dataset. Organ localization accuracy was measured using the remaining 82 volumes, which contained a total of 1504 annotated organs of which 907 were entirely contained by their volume's bounding box. Only organs that are entirely present in the volumes are used for training and test. Training and test volumes were downsampled using nearest neighbour interpolation. Integer downsampling factors were chosen so that the resulting voxel pitch was as near as possible to 3 mm per voxel in the x, y, and z directions. Downsampling to this resolution reduces memory usage without reduction in accuracy.

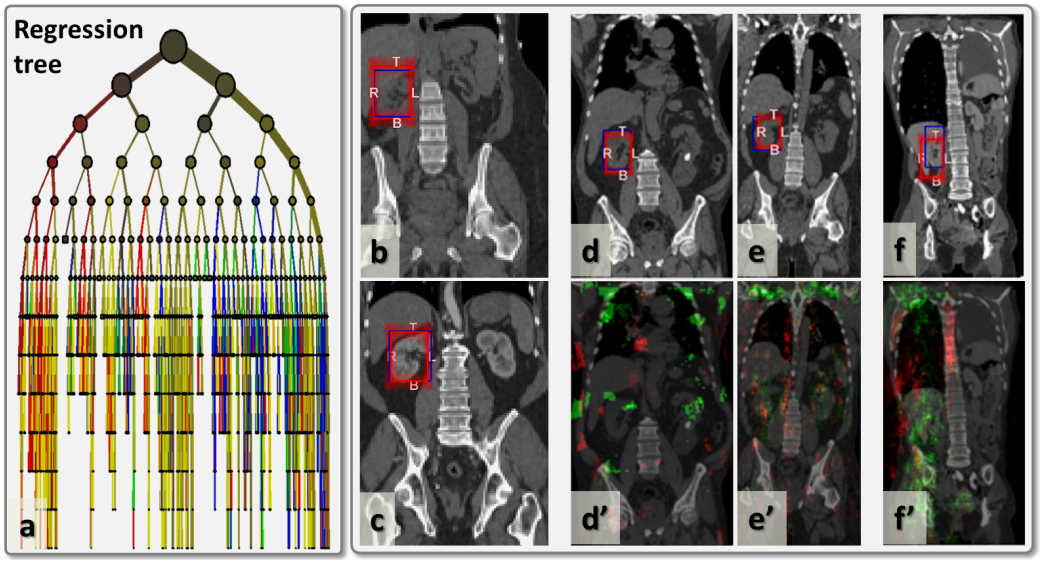


Figure 4: **Localizing the right kidney in 2D coronal images. (a)** One of the many regression trees in the forest. Different colours (red, yellow, green, blue) show the most confidently predicted box wall at each node; brighter colours indicate larger prediction confidence. **(b,c,d,e,f)** Kidney localization results in different test images. Blue is the ground-truth bounding box. The detected one is in red. Note the large anomaly in (f), a collapsed lung. **(d',e',f')** Selected landmark regions for images in (d,e,f). Red denotes landmarks which predict well the kidney position in the horizontal direction. Those are often localized along the spine, the aorta or the sides of the body. Green denotes landmark regions selected to predict the kidney's vertical position. They are localized at the top or bottom of the lungs and pelvic bones.

Quantitative evaluation. To validate the algorithm, *precision-recall* curves are plotted in fig. 5. In this context *precision* refers to the proportion of organs that were correctly detected, and *recall* to the proportion of reported detections that were correct. Precision-recall curves ares a useful means of evaluating the accuracy of detection algorithms, especially when dealing with techniques which detect different proportions of ground truth organs. Plotting how precision and recall vary as a function of a detection confidence parameter allows us to compare algorithms' accuracy at consistent recall values.

Here, a correct detection is considered to be one for which the centroid of the predicted organ bounding box is contained by the ground truth bounding

	recall level			
	0%	25%	50%	75%
precision (Our regression method)	**98%**%	**90%**	**88%**	**70%**
precision (NCC)	90%	68%	78%	45%
precision (SSD)	83%	59%	45%	34%

Table 1: **Precision-recall results for regression forest vs. template matching algorithms.** Average precision for several recall values for our regression forest method compared with template matching (using both SSD and NCC). Our method gives much higher precision at all recall levels.

box[4]. The plot shows how precision and recall vary as the detection confidence β is varied. The plot also shows a comparison with respect to template matching as described in detail later.

In fig. 5 (first row) the average precision remains high until recall reaches approximately 80%. Accuracy is best for larger organs; those with smaller size or greater positional variablity are more challenging. Selected precision/recall values are also summarized in table 1.

Table 2 shows mean localization errors for our technique, *i.e.* the absolute difference between predicted and ground truth bounding box face positions. Errors are averaged over all box faces. Despite the large variability in our test data we obtain a mean error of only 13.5 mm, easily sufficient for our intended navigation and selective retrieval applications. Errors in the axial (z) direction are approximately the same as those in x and y despite significant crop variability in this direction. Consistently good results are obtained for different choices of training set and different training runs. Notice that for partially present organs (*e.g.* cropped by the image frame) our technique still manages to find the visible box walls with decent (though sometimes slightly degraded) accuracy.

Qualitative evaluation. One application of our organ localization system is to facilitate the use of software for viewing 3D medical images by allowing the user to navigate quickly (*e.g.* with a single mouse click) to a particular anatomical structure and to have the appropriate camera parameters

[4]This definition is appropriate in light of our intended data retrieval and semantic navigation applications because the bounding box centroid would typically be used to select which coronal, axial, and sagital slices to display to the user. If the ground truth bounding box contains the centroid of the predicted bounding box, then the selected slices will intersect the organ of interest.

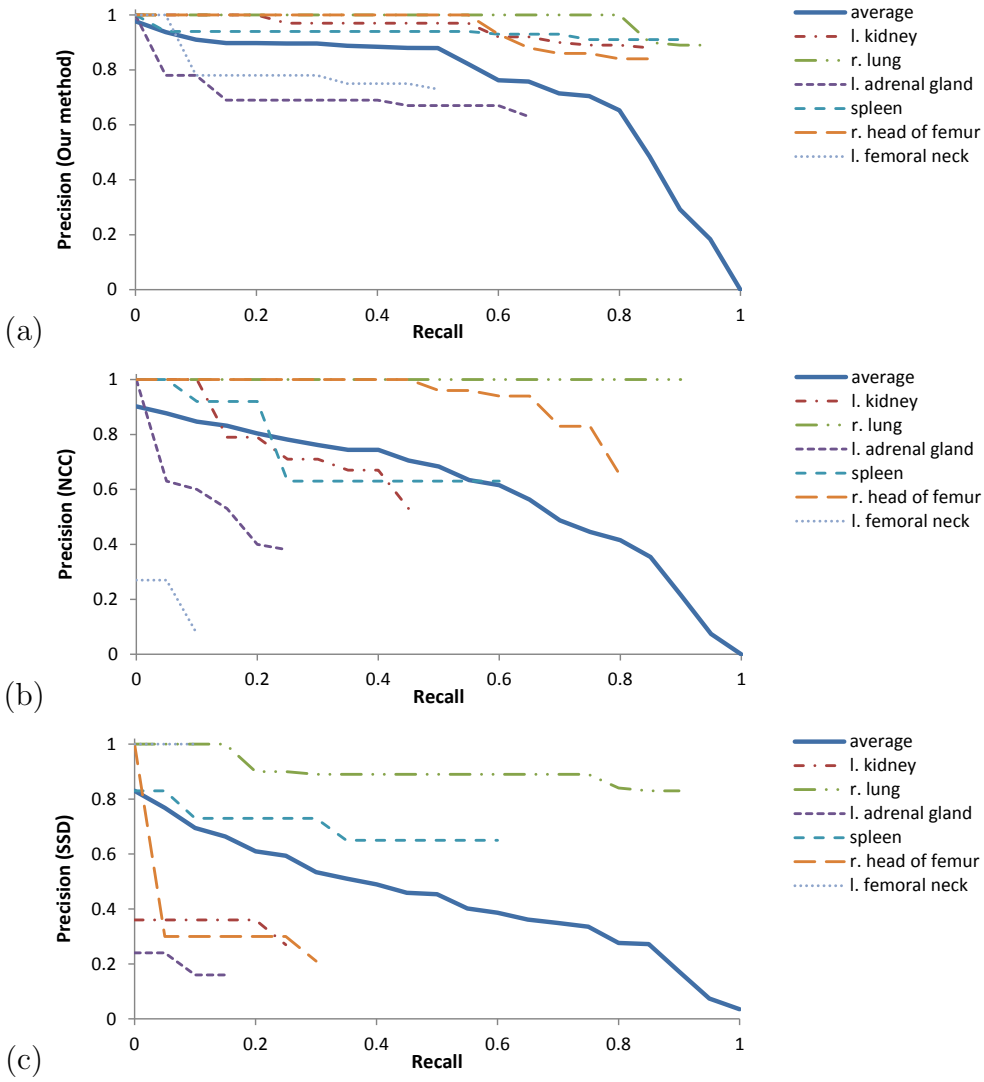


Figure 5: **Precision-recall curves for regression forests vs. template matching algorithms.** The curves show how precision and recall change as the detection confidence threshold is varied, both for a representative set of individual organ classes (some are omitted to avoid clutter) and averaged over all organ classes (solid curve). We compare results obtained using **(a)** our random regression forests technique, **(b)** template matching with the NCC distance metric, and **(c)** template matching with the SSD distance metric. In (a) the mean curve is much higher *and* the spread between different organs much less, indicating higher robustness.

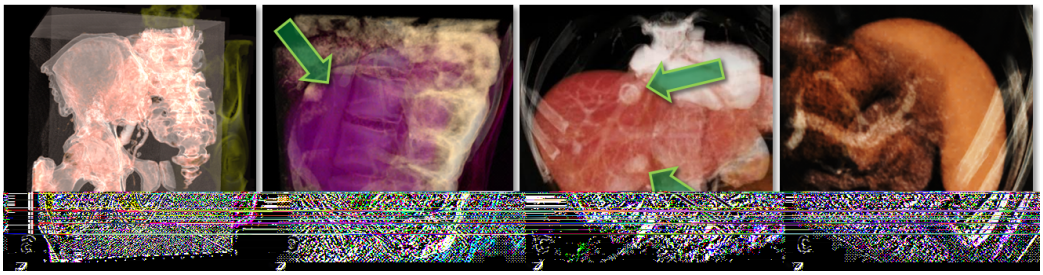


Figure 6: **Qualitative results** showing the use of our automatic anatomy localizer for semantic visual navigation within 3D renderings of large CT studies. Automatically computed bounding boxes are rendered for **(a)** a pelvic bone, **(b)** a diseased kidney, **(c)** a liver showing hemangiomas, and **(d)** a spleen. 3D camera position, cropping, and colour transfer function have been chosen automatically.

and colour transfer function selected automatically. Qualitative localization results obtained by applying the organ localization algorithm to previously unseen CT scans are shown in fig. 6.

Computational efficiency. With our C# software running in a single thread, organ detection for a typical $30 \times 30 \times 60$ cm volume requires approximately 4 s of CPU time for a typical four-tree forest. Most of the time is spent aggregating offset distributions (represented by histograms) over salient leaves. However significant speed-up could be achieved with trivial code optimizations, *e.g.* by using several cores in parallel for tree evaluation and histogram aggregation.

Comparison with template matching. We compare our method with the simple but powerful template matching algorithm. The underlying principle is to use a few selected images (templates) as exemplars for each anatomical structure. Then, given a location in a test image its classification happens via exhaustive comparison with all training exemplars for all organs (*i.e.* nearest-neighbour classification).

In order to try and capture typical CT variations we use multiple training exemplars for each organ. We manually selected five template images for each structure, from different training CT volumes so as to capture diverse appearances, sizes and shapes for the structures of interest. We obtained templates by dilating ground truth bounding boxes by 30% (this value gave the best results out of several that we tried). Now, given a test image and a specific organ class, we run a sliding-window detector approach and for

	Our method		*Elastix*		*Simplex*	
organ	mean	std	mean	std	mean	std
abdomen	**14.4**	13.4	34.6	74.2	27.6	36.5
l. adrenal gland	**11.7**	9.6	20.5	42.4	15.5	20.9
r. adrenal gland	**12.1**	9.9	22.2	45.0	18.2	29.6
l. clavicle	**19.1**	17.4	34.3	20.5	31.1	16.3
r. clavicle	**14.9**	11.6	39.0	44.3	24.1	13.9
l. femoral neck	**9.7**	7.5	38.3	78.5	16.1	15.4
r. femoral neck	**10.8**	8.3	38.4	82.3	17.3	17.7
gall bladder	**18.0**	15.0	28.1	54.5	23.2	26.6
l. head of femur	**10.6**	14.4	38.8	80.8	19.4	26.6
r. head of femur	**11.0**	15.7	39.6	84.9	19.1	28.4
heart	**13.4**	10.5	34.4	52.0	16.9	15.8
l. heart atrium	**11.5**	9.2	30.7	50.5	15.4	15.4
r. heart atrium	**12.6**	10.0	33.0	51.9	15.2	15.5
l. heart ventricle	**14.1**	12.3	35.9	51.7	18.1	16.7
r. heart ventricle	**14.9**	12.1	35.4	52.8	17.2	16.8
l. kidney	**13.6**	12.5	22.1	46.1	18.7	25.6
r. kidney	**16.1**	15.5	25.3	49.8	21.1	27.0
liver	**15.7**	14.5	26.9	53.3	23.2	30.4
l. lung	**12.9**	12.0	24.5	29.2	16.9	23.4
r. lung	**10.1**	10.1	25.0	27.2	16.0	21.7
l. scapula	**16.7**	15.7	50.9	54.1	33.1	20.1
r. scapula	**15.7**	12.0	44.4	41.2	22.7	12.4
spleen	**15.5**	14.7	29.0	46.6	23.0	22.8
stomach	**18.6**	15.8	27.6	48.9	22.8	23.4
thorax	**12.5**	11.5	36.5	37.4	25.3	35.1
thyroid gland	**11.6**	8.4	13.3	10.3	12.9	10.2
all organs	**13.5**	13.0	28.9	52.4	19.4	24.7

Table 2: **Results for forest vs. multi-atlas algorithms.** Bounding box localization errors in mm and associated standard deviations. The table compares results for our method with those for the multi-atlas Elastix- and Simplex-based registration methods. Lowest errors for each class of organ are shown in bold – our method gives lower errors for **all** organ classes.

each position we measure a similarity score with respect to all exemplars. The returned location is that for which the best similarity score is achieved. As similarity functions here we use both normalized cross-correlation (NCC) and sum of squared distances (SSD).

Comparative results are shown in fig. 5b,c and table 1. Note that to plot precision-recall curves, we need some measure of detection confidence. This stems out naturally from our probabilistic regression forest. For template matching, we use the inverse of the optimal matching cost as a proxy for detection confidence. This score is not a probability (and thus it is not comparable between organ classes). However, it does at least provide an approximately monotonic ranking, which is sufficient to plot precision-recall.

In all cases the regression forest approach produces much higher precision at all recall levels. Furthermore, the spread between the curves associated with different organs is smaller in the forest case, indicating higher robustness. Template matching works very well for a few visually distinct organs such as the lungs, but much worse for structures with a higher variability in shape or appearance (*e.g.* the spleen). Normalized cross correlation performs slightly better than SSD in this regard, but both approaches give significantly lower precision than our algorithm at all recall levels. One of the main reasons seems to be that despite the use of multiple exemplars template matching is not robust enough to deal with the high variability (in appearance, location, size and shape) observed in our dataset. Using a larger number of templates should help improve the accuracy, but at a higher computational cost. This suggests better *generalization* behavior for forests compared to template matching.

Further discussion on regression vs. detection. In Zhou et al. (2007) the authors report a comparison between boosted regression and an anatomy detector.[5] They show superiority of regression in terms of efficiency but no significant improvement in terms of precision.

Figure 5 has already demonstrated how our regression forest produces higher localization accuracy than a carefully constructed template-based anatomy detector. Additionally, we have compared our bounding-box localization results with those obtained by the voxel-wise classification forest in Criminisi et al. (2009) and have found that regression forests achieve errors which are less than half of those obtained with classification.[6] In addition, our regression-based parametrization gives an indication of organ extent as well as its position.

Finally, the classification approach in Liu et al. (2010) is very different from ours and not directly comparable. In Liu et al. (2010) the authors train a *supervised* classifier to be able to localize *landmark* points. To do so they need labelled training landmarks. In contrast, in our work a supervised regressor is trained on ground-truth bounding boxes. In our case the goal is to optimize the organ (and not the landmark) localization. Discriminative, intermediate landmark regions are also obtained, but as a by-product and

[5] In Zhou et al. (2007) a probabilistic boosting tree-based detector is employed.

[6] Comparative experiments were run on exactly the same randomly selected training and test data.

without supervision; *i.e.* we do not need labelled training landmark points.

Comparison with affine, atlas-based registration. Yet another popular strategy for anatomy localization is to align the input volume with a suitable *atlas*, *i.e.* a reference scan for which organ bounding box positions are known. Bounding box positions are then determined by using the estimated geometric transform to map box locations from the atlas into the input image.

Non-linear atlas registration (via non-rigid registration algorithms) can, in theory, provide accurate localization results. In practice however, this approach is sensitive to bad initialization and requires significantly greater computation times than our regression approach. Since speed is an important aspect of our work, here we chose to compare our results with those from a comparably fast atlas-based algorithm based on *global affine registration*. This is a rather approximate approach because accuracy is limited by inter- and intra-subject variability in organ location and size. However, it is robust and its computation times are closer to those of our method.

A multi-atlas approach is used here to try and capture data variability (Isgum et al., 2009). From the training set, five scans were selected to be used as atlases. The selected scans included three abdominal-thorax scans (one female, one male and one slightly overweight male), one thorax scan, and one whole body scan. This selection was representative of the overall distribution of image types in the database. All five atlases were registered to all the scans in the test set. For each test scan, the atlas that yielded the smallest registration cost was selected as the best one to represent that particular test scan. Registration was achieved using two different global affine registration algorithms. The first algorithm ('Elastix') is that implemented by the popular *Elastix* toolbox (Klein et al., 2010) and works by maximizing mutual information using stochastic gradient descent. The second algorithm ('Simplex') is our own implementation and works by maximizing correlation-coefficient between the aligned images using the simplex method as the optimizer (Nelder and Mead, 1965).

Resulting errors (computed on the same test set) are reported in table 2. The atlas registration techniques give larger mean errors and error standard deviation (nearly double in the case of Elastix) compared to our approach. Furthermore, atlas registration requires between 90 s and 180 s per scan (*cf.* our algorithm runtime is ~ 4 s for $T = 4$ trees).

Figure 7 further illustrates the difference in accuracy between the approaches. For the atlas registration algorithms, the error distribution's larger

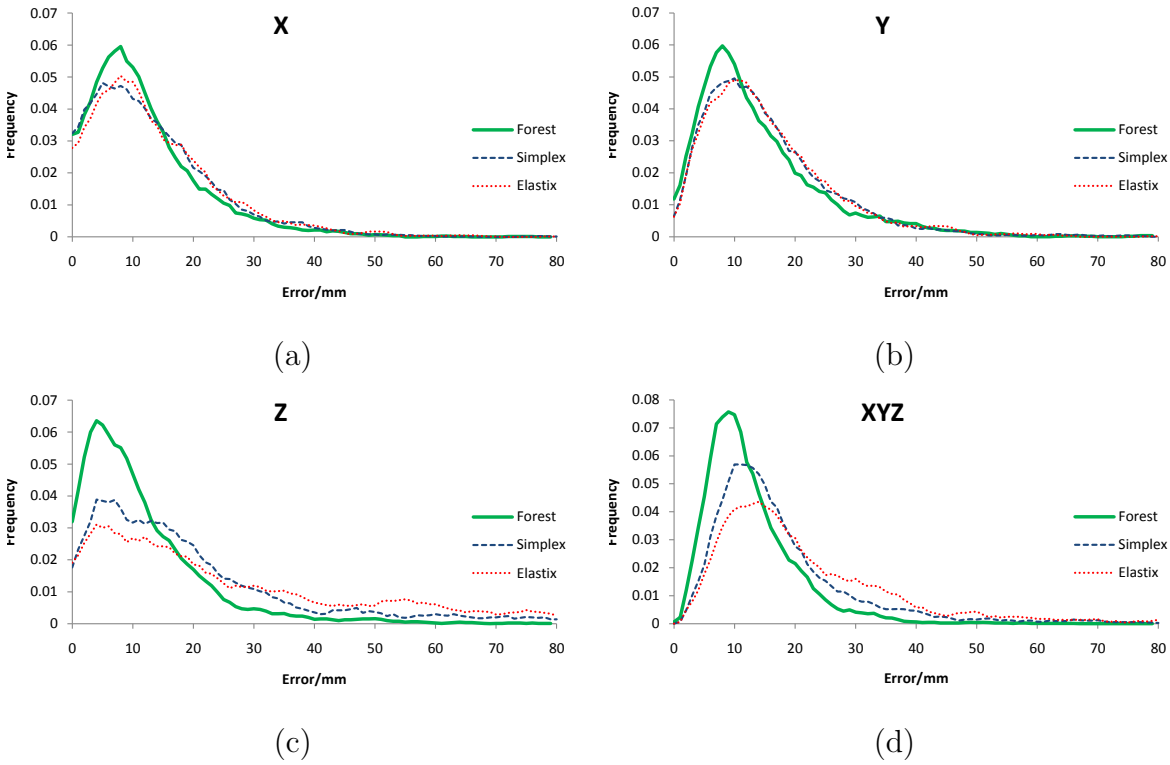
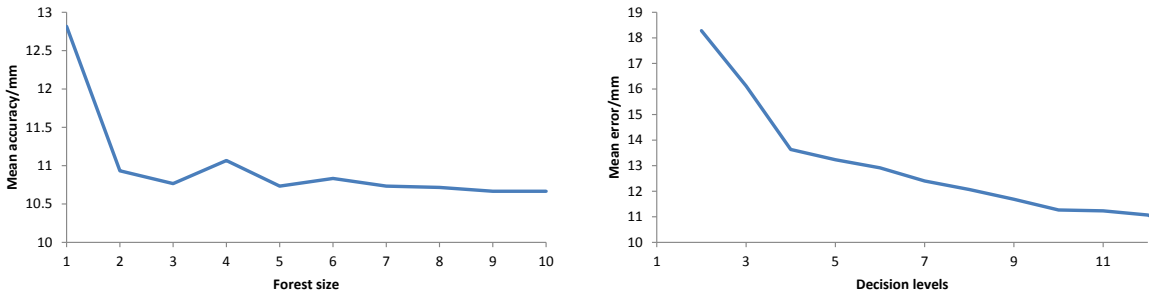


Figure 7: **Prediction errors for forest vs. multi-atlas algorithms.** Distributions of bounding box localization errors for our algorithm ('Forest') and two atlas-based techniques ('Elastix' and 'Simplex'). Error distributions are shown separately for **(a)** left and right, **(b)** anterior and posterior, and **(c)** head and foot faces of the detected bounding boxes, and **(d)** averaged over all bounding box faces for each organ. The error distributions for the atlas techniques (particularly in plots (c) and (d)), have more probability mass in the tails, which is reflected by larger mean errors and error standard deviations.

tails suggest a less robust behavior[7]. This is reflected in larger values of the error mean and standard deviation and is consistent with our visual inspection of the registrations. In fact, in about 30% of cases the registration process got trapped in local minima and produced grossly inaccurate alignment. In those cases, results tend not to be improved by using a non-linear registration step (which tends not to help the registration algorithm to escape bad local minima, whilst increasing the runtime considerably).

[7]Because larger errors are produced more often than in our algorithm.



(a) (b)

Figure 8: **Mean error in mm** (a) with fixed tree depth $D = 12$ and varying forest size T and (b) with varying maximum tree depth D and fixed forest size $T = 4$. Errors (averaged over six bounding box faces) were computed on previously unseen test scans. To ensure error statistics remain comparable as parameters T and D vary, the detection confidence threshold β was tuned for each parameter setting to give recall of (a) 0.8 and (b) 0.5.

Computational and memory efficiency. A regression forest with 4 trees and 12 decision levels requires $\sim$ 10MB of memory, which compares very favourably with the roughly 100MB required for each atlas. Furthermore, the problems of model size and runtime performance are exacerbated by the use of more accurate and costly multi-atlas techniques (Isgum et al., 2009). In our algorithm increasing the size of the training set usually decreases the test error without significantly affecting the test runtime, whereas with multi-atlas techniques increasing the number of atlases linearly increases the runtime.

Accuracy as a function of forest parameters. Fig. 8 shows the effect of tree depth and forest size on the accuracy of bounding box predictions. Accuracy improves with tree depth up to around 12 levels. As expected increasing the forest size T produces monotonic improvement without significant overfitting. Good performance is obtained with as few as two or three trees.

Automatic landmark detection. Fig 9 shows how we can visualise the anatomical landmark regions that were selected automatically for organ localization during regression tree training. Given a trained regression tree and an input volume, we select one or two leaf nodes with high prediction confidence for a chosen organ class (*e.g.* `l. kidney`). Then, for each sample arriving at the selected leaf nodes, we shade in green the cuboidal regions of the input volume that were used during evaluation of the parent nodes' feature tests.

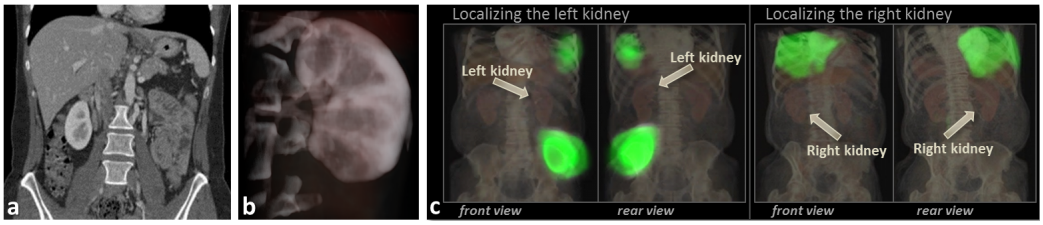


Figure 9: **Automatic discovery of salient anatomical landmark regions. (a)** A test volume and **(b)** a 3D volume rendering of the left kidney's bounding box, as detected by our algorithm. **(c)** The highlighted green regions correspond to regions of the volume that were automatically selected as salient predictors of the position of the kidneys.

Thus, the green regions represent some of the anatomical locations that were used to estimate the location of the chosen organ. In this example, the bottom of the left lung and the top of the left pelvis are used to predict the position of the left kidney. Similarly, the bottom of the right lung is used to localize the right kidney. Such regions correspond to meaningful, visually distinct, anatomical landmarks that have been computed without any manual tagging.

Finally, notice that voxels within the organ itself do not contribute much to its localization. This is in accordance with other landmark-based approaches; but here informative landmarks have been selected completely automatically.

Robustness with respect to field of view. Our training database contains many, very diverse scans. In particular they are very different from one another in their field of view, with some presenting very cropped views of *e.g.* the abdomen and the organs within. When tested using partially visible organs we found that our system still worked correctly. However, in those cases only visible landmarks contribute to the organ localization. As a consequence, the localization of cropped organs tends to be associated with a lower confidence in the output posterior, as expected.

4. Conclusion

Anatomy localization has been cast here as a non-linear regression problem where *all* voxel samples vote for the position of all anatomical structures. Location estimates are obtained by a multivariate regression forest algorithm

that is shown to be more accurate and efficient than competing registration-based and template-matching techniques.

At the core of the algorithm is a new information-theoretic metric for regression tree learning which works by maximizing the confidence of the predictions over the position of all organs of interest, simultaneously. Such strategy produces accurate predictions as well as meaningful anatomical landmark regions.

Accuracy and efficiency have been assessed on a database of 400 diverse CT studies. The usefulness of our algorithm has already been demonstrated in the context of systems for efficient visual navigation of 3D CT studies (Pathak et al., 2011) and robust linear registration (Konukoglu et al., 2011). Future work will include extension to imaging modalities other than CT and the exploration of different context models.

Appendix – Background on regression forests

Regression trees (Breiman et al., 1984) are an efficient way of mapping a complex input space to continuous output parameters. Highly non-linear mappings are handled by splitting the original problem into a set of smaller problems which can be addressed with simple predictors.

Figure 10 shows illustrative 1D examples where the goal is to learn an analytical function to predict the real-valued output y (*e.g.* house prices) given the input x (*e.g.* air pollution). Learning is supervised as we are given a set of training pairs (x, y). Each node in the tree is designed to split the data so as to form clusters where accurate prediction can be performed with simpler models (*e.g.* linear in this example). More formally, each node performs the test $\xi > f(x) > \tau$, with ξ, τ scalars. Based on the result each data point is sent to the left or right child.

During training, each node test (*e.g.* its parameters ξ, τ) is optimized so as to obtain the best split; *i.e.* the split that produces the maximum reduction in geometric error. The error reduction r is defined here as: $r = e(\mathcal{S}) - \sum_{\mathtt{i} \in \mathtt{L},\mathtt{R}} \omega_i e(\mathcal{S}_i)$ where $\mathcal{S}$ indicates the set of points reaching a node, and `L` and `R` denote the left and right children (for binary trees). For a set $\mathcal{S}$ of points the error of geometric fit is: $e(\mathcal{S}) = \sum_{j \in \mathcal{S}} [y_j - y(x_j; \boldsymbol{\eta}_S)]^2$, with $\boldsymbol{\eta}_S$ the two line parameters computed from all points in $\mathcal{S}$ (*e.g.* via least squares or RANSAC). Each leaf stores the continuous parameters $\boldsymbol{\eta}_S$ characterizing each linear regressor. More tree levels yield smaller clusters and smaller fit errors, but at the risk of overfitting (Criminisi and Shotton, 2013).

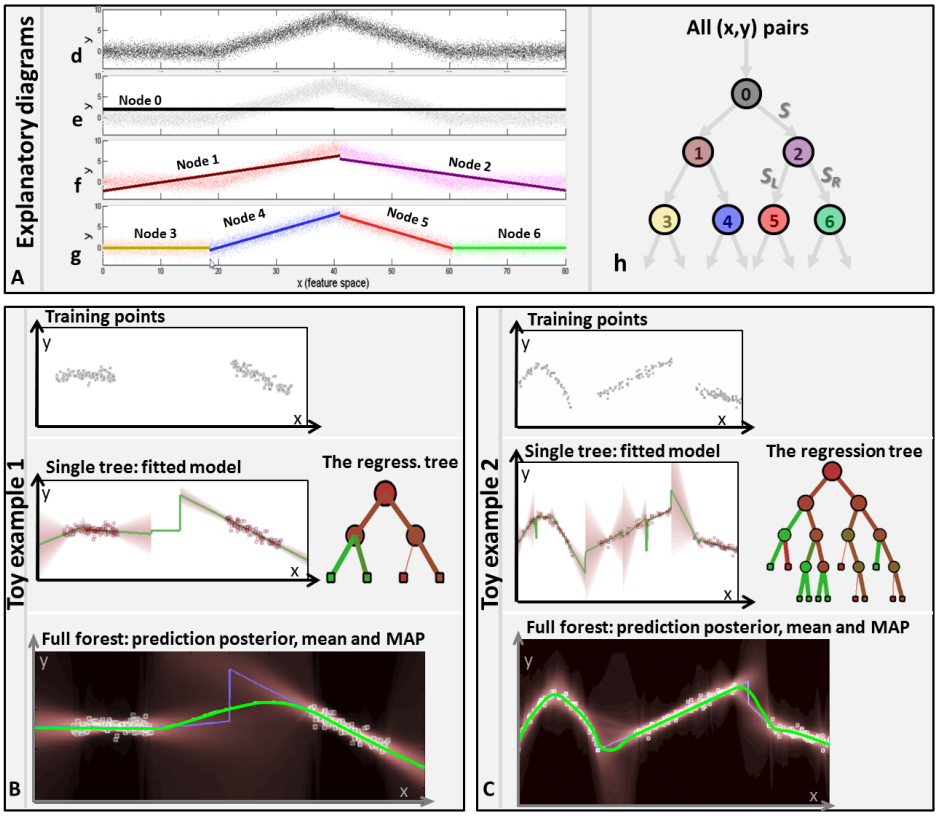


Figure 10: **Regression forest:** explanatory 1D examples. **(panel A)** (d) Training data points. (e) A single linear function fits the data badly. (f, g) Using more tree levels yields more accurate fit of the regressed model. Complex non-linear mappings are modeled via a hierarchical combination of many, simple linear regressors. (h) The corresponding regression tree. **(panel B, C)** Further examples. Training points are shown with grey squares. Each regression tree fits a piece-wise linear model to the data. At each leaf, each line has an associated uncertainty. Testing the forest (bottom row) corresponds to computing the distribution $p(y|x)$ at each x value. The conditional mean $E[y|x]$ (in green) smoothly interpolates the training points in large gaps. Away from training points the uncertainty increases, as expected.

References

Barbu, A., Suehling, M., Xu, X., Liu, D., Zhou, S.K., D., C., 2012. Automatic detection and segmentation of lymph nodes from ct data. IEEE Trans. Medical Imaging 31, 240–250.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and Regression Trees. Chapman and Hall/CRC.

Criminisi, A., Shotton, J., 2013. Decision Forests for Computer Vision and Medical Image Analysis. Advances in Computer Vision and Pattern Recognition, Springer.

Criminisi, A., Shotton, J., Bucciarelli, S., 2009. Decision forests with long-range spatial context for organ localization in CT volumes, in: MICCAI workshop on Probabilistic Models for Medical Image Analysis.

Fenchel, M., Thesen, S., Schilling, A., 2008. Automatic labeling of anatomical structures in MR fastview images using a statistical atlas, in: MICCAI.

Feulner, J., Zhou, S.K., Seifert, S., Cavallaro, A., Hornegger, J., Comaniciu, D., 2009. Estimating the Body Portion of CT Volumes by Matching Histograms of Visual Words, in: Pluim, J.P.W., Dawant, B.M. (Eds.), Proceedings of SPIE.

Gall, J., Lempitsky, V., 2009. Class-specific Hough forest for object detection, in: IEEE CVPR, Miami.

Gueld, M.O., Kohnen, M., Keysers, D., Schubert, H., Wein, B.B., Bredno, J., Lehmann, T.M., 2002. Quality of DICOM header information for image categorization, in: SPIE Storage and Retrieval for Image and Video Databases.

Hardle, W., 1990. Applied non-parametric regression. Cambridge University Press.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. PAMI 20.

Isgum, I., Staring, M., Rutten, A., Prokop, M., Viergever, M.A., van Ginneken, B., 2009. Multi-atlas-based segmentation with local decision fusion-application to cardiac and aortic segmentation in ct scans. IEEE Trans. Medical Imaging 28.

Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P., 2010. elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imaging 29, 196–205.

Konukoglu, E., Criminisi, A., Pathak, S., Robertson, D., White, S., Haynor, D., Siddiqui, K., 2011. Robust linear registration of ct images using random regression forests, in: SPIE Medical Imaging, Orlando, Florida, US.

Liu, D., Zhou, K., Bernhardt, D., Comaniciu, D., 2010. Search strategies for multiple landmark detection by submodular maximization, in: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pp. 2831–2838.

Montillo, A., Ling, H., 2009. Age regression from faces using random forests, in: ICIP.

Nelder, J., Mead, R., 1965. A simplex method for function minimization. The computer journal 7, 308.

Pathak, D.S., Criminisi, A., Shotton, J., White, S., Robertson, D., Sparks, B., Munasinghe, I., Siddiqui, K., 2011. Validating automatic semantic annotation of anatomy in dicom ct images, in: SPIE Medical Imaging, Orlando, Florida, US.

Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Sühling, M., Cavallaro, A., Comaniciu, D., 2009. Hierarchical parsing and semantic navigation of full body CT data, in: Pluim, J.P.W., Dawant, B.M. (Eds.), SPIE.

Shimizu, A., Ohno, R., Ikegami, T., Kobatake, H., 2006. Multi-organ segmentation in three-dimensional abdominal CT images. Int. J CARS 1.

Shotton, J., Winn, J., Rother, C., Criminisi, A., 2009. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout and context., in: IJCV.

Torralba, A., Murphy, K.P., Freeman, W.T., 2007. Sharing visual features for multiclass and multiview object detection. IEEE Trans. PAMI .

Yao, C., Wada, T., Shimizu, A., Kobatake, H., Nawano, S., 2006. Simultaneous location detection of multi-organ by atlas-guided eigen-organ method in volumetric medical images. Int. J CARS 1.

Yin, P., Criminisi, A., Essa, I., Winn, J., 2007. Tree-based classifiers for bilayer video segmentation, in: CVPR.

Zhan, Y., Zhou, X.S., Peng, Z., Krishnan, A., 2008. Active scheduling of organ detection and segmentation in whole-body medical images, in: MICCAI.

Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D., 2007. Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features, in: ICCV.

Zheng, Y., Georgescu, B., Comaniciu, D., 2009a. Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images, in: IPMI '09: Proc. of the 21st Intl Conference on Information Processing in Medical Imaging.

Zheng, Y., Lu, X., Georgescu, B., Littmann, A., Mueller, E., Comaniciu, D., 2009b. Robust object detection using marginal space learning and ranking-based multi-detector aggregation: Application to left ventricle detection in 2D MRI images, in: CVPR.

Zhou, S., Georgescu, B., Zhou, X., Comaniciu, D., 2005. Image-based regression using boosting method, in: ICCV.

Zhou, S.K., Zhou, J., Comaniciu, D., 2007. A boosting regression approach to medical anatomy detection. IEEE CVPR 0, 1–8.