VISCERAL Detection2 Benchmark

Guidelines for Participation v1.0 (20150619)

Document History

v1.0 - 20150619 - Released version of document

1. Introduction

1.1 Aim of the Benchmark

The aim of the benchmark is to detect all lesions in the provided image volumes. Each lesion in a volume should be detected and marked by one point, irrespective of how large the lesion is. If two points are located in a large lesion, then one of the detected points will be a false positive.

1.2 Registration

The first step in participation is registration. This is done online on the following page:
http://visceral.eu:8080/register/Registration.xhtml
Select the “Detection2” benchmark on the registration screen.
During the registration process, participants will be required to sign and upload a participation agreement. Once the participant is registered and the participation agreement has been accepted by the organisers, the account will be activated.
Once the account is activated, logging into the registration system will reveal the *participant dashboard*.

1.3 Participant Dashboard

Virtual Machines are not provided for this Benchmark, so all information about them should be ignored on the DashBoard. The main purpose of the Dashboard is to serve as a repository for documents about the Benchmark.

Documents about the Benchmark, including a list of volumes can be found by pressing 

. The evaluation software that will be used for calculating the evaluation metrics can be downloaded (see Section 6).

1.4 Data

Instructions for downloading the data via FTP are provided in a document found by pressing 

.

A list of all files in the training data set can be downloaded from the participant dashboard in the registration system by clicking on 

.

2. Benchmark Organisation

2.1 Resources provided and their use

Participants should download:

- **The training and testing data** consisting of imaging data and corresponding annotations of lesions in the imaging data (lesion annotation consists of the center coordinates for small lesions, or of center and 2 boundary points for larger lesions)
- The **evaluation metric calculation software** (see Section 6).
- The **organ masks** consisting of binary masks that indicate the organs in which lesion detection is performed and evaluated.

Detector training, parameter tuning, and repeated experiments should be done **only on the training data**. Participants should test the algorithms on the test data only using the provided metric calculation program.

2.2 Result publication

When results obtained using the VISCERAL Detection2 Benchmark Resources are published, please do the following:

- Link the publication from http://bibsonomy.org with the tag *visceral-detection2*.
- Reference the following paper:
  Georg Langs, Henning Müller, Bjoern H. Menze and Allan Hanbury, VISCERAL: towards large data in medical imaging - challenges and directions. Proc. MICCAI 2012 Workshop on Medical Content-based Retrieval for Clinical Decision Support (MCBR-CDS), 2012, Springer LNCS 7723, pages 92–98, Nice, France.

3. Further Information

All participants and organisers are automatically registered to the participants-detection2@visceral.eu mailing list, and can post on the mailing list. Use this list to communicate only among participants and the organisers, to ask questions, draw attention to problems or share hints and tips.

A LinkedIn group has been set-up for discussion about the Benchmark. Ask questions and make comments on this group:
http://www.linkedin.com/groups/VISCERAL-Benchmark-Discussion-5089631

4. Evaluation

4.1 Detection metrics

In the detection task, an annotated lesion, L, is represented by three points, namely the center of the lesion, C_i, and two other points, $D1_i$ and $D2_i$, indicating the diameter of the

lesion. Participating algorithms are expected to provide per lesion exactly one point, P_i, as near as possible to the center of the lesion, C_i.

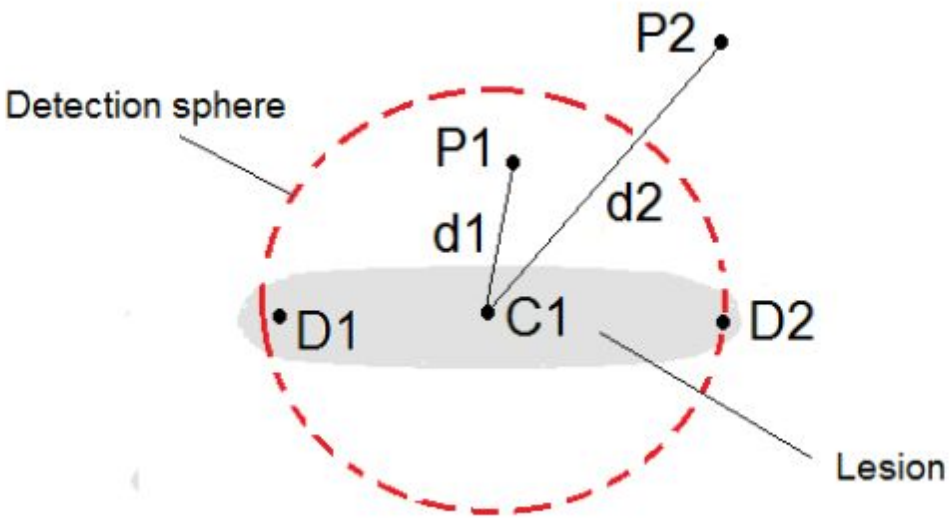


Figure: Schematic representation of a lesion annotated by the center C1 and the two diameter points D1 and D2. The points P1 and P2 are retrieved by an algorithm. P1 lies within the detection sphere and is thus considered as detected in contrast of the point P2.

The evaluation of the detection task takes place at three different levels:

1. **Lesion level**: for each annotated lesion, two values are measured, namely
    - Minimum Euclidean distance, min(d_i): for each annotated lesion, the distance to the nearest point retrieved by the participating algorithm is measured as shown in the Figure. This distance is provided for each annotated lesion, regardless of whether the lesion is considered as detected or not.
    - Detection: a lesion is considered as detected if the point P_i, provided by the algorithm, is within the imaginary sphere centered on C_i and has the diameter given by the points D1 and D2. In particular, a radius of the sphere, r, is considered, which is equal to the distance between the center C_i and the farthest of the points D1 and D2. That is, a lesion is detected iff min(d) < r. In the Figure, the point P1 is detected and P2 is not detected.
2. **Volume level**: The confusion matrix (true positives, false positives, true negatives and false negatives) is calculated per volume, based on the detection values calculated in (i). From this confusion matrix, the precision (Percentage of correctly detected lesions), and the recall (Percentage of total lesions detected) are calculated for each volume and participating algorithm. As it is expected that algorithms provide exactly one point per lesion, all other points that may be retrieved are considered as false positives.
3. **Structure average level**: To test whether the scores of lesion detection are generally dependant on the structure, we calculate score averages (the Euclidean distance) for each structure over all volumes/participants.

4.2 Organ Masks to Facilitate Detection and Evaluation

As mentioned above, it is expected that exactly one point per lesion is retrieved by each participating algorithm. To penalize algorithms that may try to improve evaluation results by providing many points, all other points retrieved are considered as false positives. However, annotators have looked at specific regions of the volume, which means that one cannot be sure that other regions are free of lesions. In other words, participating algorithms could detect lesions that have not been annotated. To avoid penalizing such lesions, binary masks are used for each volume, which mask only those regions that have been annotated. Retrieved points that lie outside the mask are not considered in the confusion matrix. These masks are also provided to the participants.

5. File Format Conventions

The following file format should be used for the results of lesion detection, as it is required by the evaluation metric calculation software:

PatientID_ModalityCounter_ModalityName_Lesions.fcsv

```
#
# Each row entry in this file contains a lesion location for the respective
volume
# An entry is structured as follows lesion, <x-coordinate>, <y-coordinate>,
<z-coordinate>
# The coordinates are provided in world (scanner) coordinates
#
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
lesion,x,y,z
<...>
```

6. Evaluation Software

The EvaluateSegmentation software is used to calculate the evaluation metrics. The latest version of the EvaluateSegmentation software is always available here: https://github.com/codalab/EvaluateSegmentation

The syntax for using the EvaluateSegmentation tool to evaluate the detection task is the following:

EvaluateSegmentation -det groundTruthPath testPath [-mask maskPath] [-xml xmlPath]

where
groundTruthPath: the absolute path (or complete url) to the file containing the ground truth lesion positions.
testPath: path (or complete url) to the file containing the lesions detected automatically.
maskPath: path (or complete url) to optional image to be used as a region of interest (mask).
xmlPath: name of xmlFile to additionally save the detection evaluation results (optional).

Example:

EvaluateSegmentation -det groundtruth.fcsv test.fcsv -mask mask.nii -xml results.xml

This command compares the detected lesions in the file test.fcsv with the ground truth lesions in groundtruth.fcsv using the image mask.nii as an ROI and saves the results in the file results.xml in addition to displaying them on the screen.

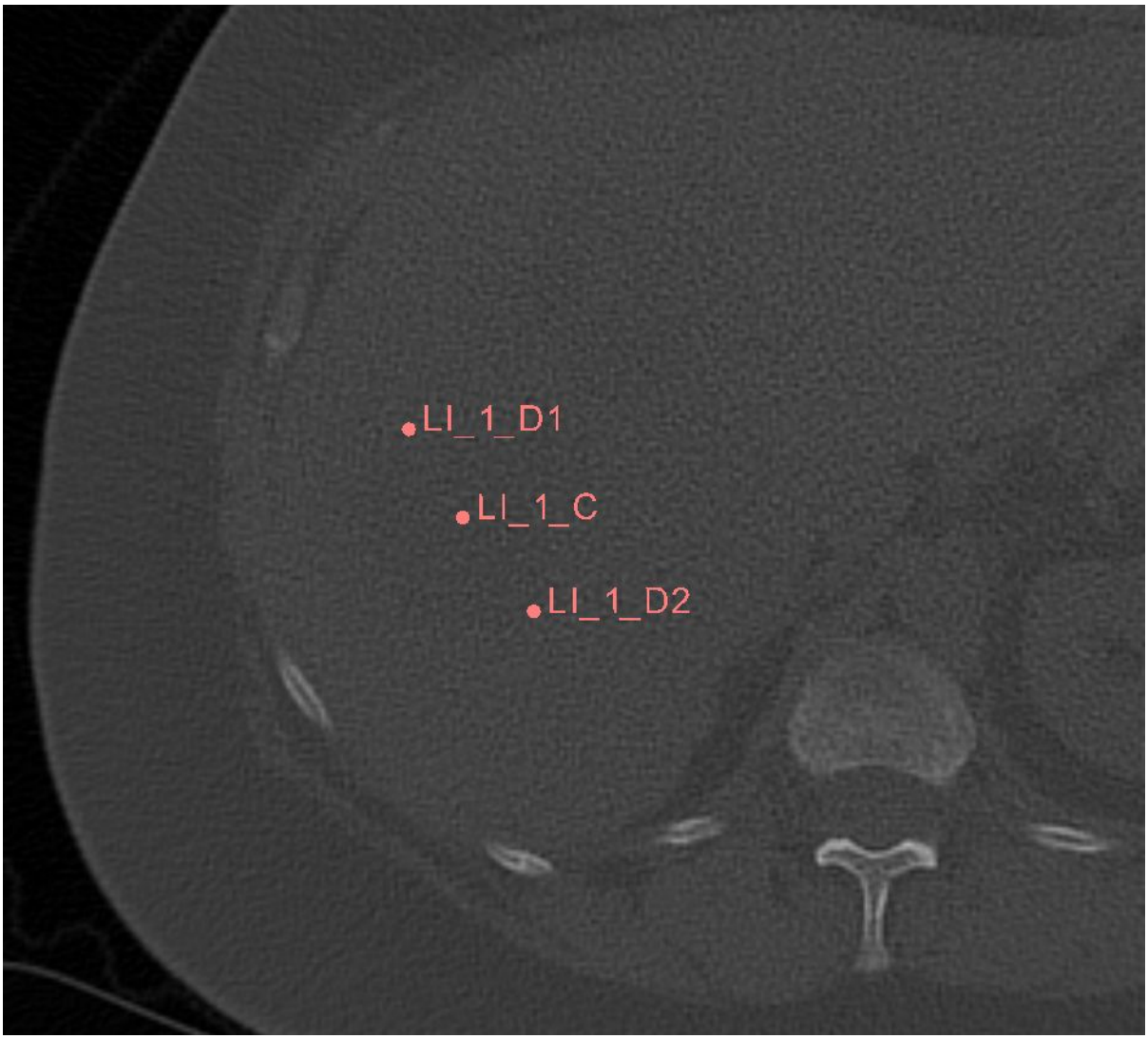
Note: the new version of EvaluateSegmentation that supports the feature above (evaluation of lesion detection) will be available in the first week of July for download via GitHub.

7. FAQ

- For the lesion annotation (see pic. below) I concluded that **C** - stands for *centre*. While **D** - stands for *diameters*. Is this correct?

C is always the center, in case of large lesions that annotators were instructed to annotate two additional points on the perimeter, to give an estimate of the radius. Since the lesions are not spherical, this is an estimate, but in this context it is still clinically relevant.

- For subject 10000018, there might be some mistake regarding the liver lesion annotation of this patient (see image below).



It is a lesion, but it is poorly visible in CT since there was no contrast agent applied. It might get better when playing with the contrast, but likely not a lot.

Specifically it is a set of multiple confluenting haemangiomas, i.e., a large lesion which almost occupies the whole right liver lobe and is inhomogeneous with different gray shades.