VISCERAL Detection Benchmark

Guidelines for Participation v1.2 (20141118)

Document History

v1.0 - 20140922 - Released version of document
v1.1 - 20141030 - Added an FAQ section at the end
v1.2 - 20141118 - Updated section 4 on the metrics to be used for evaluation

1. Introduction

1.1 Aim of the Benchmark

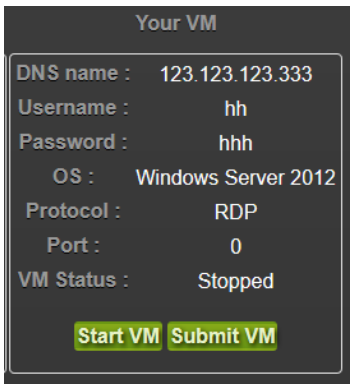
The aim of the benchmark is to detect all lesions in the provided volumes. Each lesion in a volume should be detected and marked by one point, irrespective of how large the lesion is. If two points are located in a large lesion, then one of the detected points will be a false positive.

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
During the registration process, participants will be required to sign and upload a participation agreement.
Once the participant is registered, logging into the registration system will reveal the *participant dashboard*.

1.3 Virtual Machine

The medical imaging data is stored on the Microsoft Azure Cloud. When participants register successfully, they will receive a virtual machine (VM) in the Microsoft Azure cloud (Windows or Linux VMs are available), provided and financed by VISCERAL, with the support of Microsoft Research. The information for accessing the VM will appear on the participant dashboard in the registration system (after a delay of up to a week).

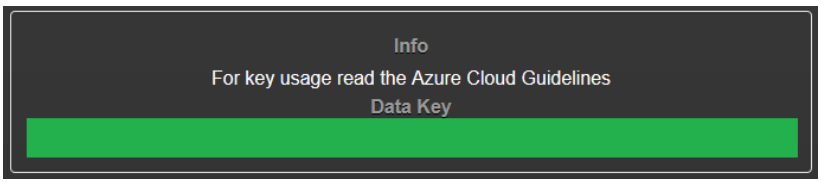


Note that each participating group should only register for the VISCERAL Benchmark once. After successful registration, the person completing the registration will get root access to the assigned VM, and will be able to create logins for colleagues.
Please shut down the VM if it will not be used for a longer time (using the “Stop VM” button on the participant dashboard in the registration system). It can again be started using the “Start VM” button.

Documents about the Benchmark, including information on using the VM and a list of volumes can be found by pressing View Benchmark Files. The evaluation software that will be used for calculating the evaluation metrics can be downloaded (see Section 6).

1.4 Training Data

The training data can be accessed from the VM. The *Data Key* for accessing the data is provided on the participant dashboard once the VM is assigned. Please only access the data on the cloud from within the assigned VM — accessing the data from outside the cloud results in additional costs for the organisers. Instructions for downloading the data via FTP are provided in a document found by pressing View Benchmark Files.



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

The image file URLs are constructed as:

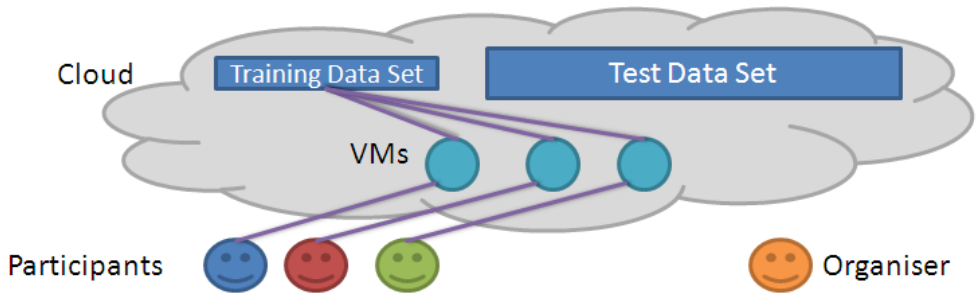
`cURL+filename+saKey`

cURL: container URL,
`http://visceralstorage1.blob.core.windows.net/trainingset/`
image filename: `PatientID_ModalityCounter_ModalityName.nii.gz`
lesion annotation filename: `PatientID_ModalityCounter_ModalityName_Lesions.fcsv`
saKey: shared access key, e.g.
`?sr=c&si=readonly&sig=Z69O9Vz8TU0RxawtASpmpWZnT%2FhF2OgJOI7iEt60mis%3D`

2. Benchmark Organisation

The benchmark runs in two phases.

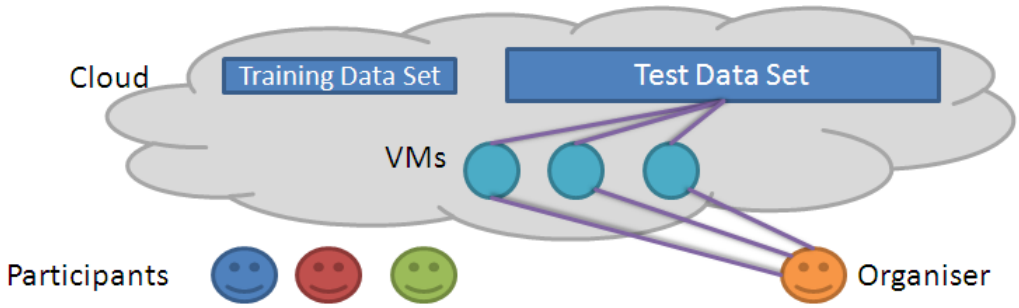
2.1 Training phase



The participants each have their own VM in the cloud, linked to an annotated training data set of the same structure as the test data set. Software for carrying out the benchmark tasks must be placed into the VMs by the participants by the submission deadline. The

software must be at least executable binaries and all libraries and other support required to execute the software. Source code is not required and must be removed from the VM by the submission deadline if the organizers should not see it (although even if source code is there, the organisers will not copy it to anywhere outside of the VM that it is in. Participants requiring additional security with respect to code or binaries can request to sign an NDA with the organisers). The software must satisfy all specifications in this document. The test data set is not accessible to the participants.

2.2 Testing phase



On or before the benchmark submission deadline, when all software is installed on the VM as required by the specifications, **participants must press** Submit VM **on the participant dashboard** to be included in the testing phase. This lets the organizers know that they can take over the VM from the participants, execute the software installed on the VMs on the unseen test data set and evaluate the results. For all participants, access to the VMs will be stopped after the submission deadline, but evaluation will only be done for the VMs for which the "Submit VM" button was pressed. The Virtual Machines will be monitored regularly during the evaluation phase. Delays caused by the participants executable (crashes, error messages ...) during the automatic evaluation phase might influence the total number of volumes evaluated. Participants should also make sure that the output segmentations and landmark files are compatible with the evaluation tool before submitting their VM.

3. Further Information

All participants and organisers are automatically registered to the Participants-VISCERALdetection@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

You can also follow VISCERAL on Twitter: @VisceralEU.

4. Evaluation metrics

To evaluate the detection task, three metrics will be used, namely the precision (Percentage of correctly detected lesions), the recall (Percentage of total lesions detected), and the Euclidean distance. The calculation of these metrics is described in the following paragraphs.

For each lesion, an algorithm should provide only one point (c) representing the center of the lesion. The evaluation is based on comparing this point with the corresponding ground truth, which consists of three points, one representing the center (C), and two other points (D1, D2) representing the diameter. A lesion is considered as detected if the point c provided by the algorithm is within the imaginary sphere centered on C and has the diameter given by D1 and D2. The confusion matrix (true positives, false positives, true negatives and false negatives) is calculated per volume, from which the precision and the recall are calculated per volume.

Note that only one point should be provided per lesion, which means that each other point will be considered as false positive. In some cases where unregistered lesions could exist outside of the test region, binary masks of the tested organ will be used, where only points within this mask will be considered and all other points are ignored.

The third metric (distance) is given as the Euclidean distance between the point provided by the algorithm (c) and the center of the lesion (C). Furthermore, we will calculate for each lesion a specific average error, i.e. the Euclidean distance averaged over all volumes.

5. Program and File Format Conventions

After the submission deadline, for those VMs submitted by participants (“Submit VM” button), the organisers will run the participants’ programs installed in the VMs on the unseen test data. Please ensure that all of the following naming and calling conventions and file format conventions are followed, to ensure that this works smoothly.

5.1 Program naming and Calling

One executable file (can be a script or compiled program) with the name

`execute_standard` or `execute_standard.{extension}`

must be in the home directory of the *azureuser* user (for those using Linux VMs), or on the Windows Desktop of the *azureuser* user (for those using Windows VMs) of the virtual machine. The *azureuser* user will be the only username available when the participant is assigned the virtual machine after registration.

Parameters

This executable file must take the following set of parameters, which are explained below:

```
-i [URL file to segment]
```

```
-o [output path]
-c [ConfigurationID]
```

The image file URLs are constructed as:

```
cURL+filename+saKey
```

cURL: container URL,
`http://visceralstorage1.blob.core.windows.net/trainingset/`
filename: `PatientID_ModalityCounter_ModalityName.nii.gz`
saKey: shared access key, e.g.
`?sr=c&si=readonly&sig=Z69O9Vz8TU0RxawtASpmpWZnT%2FhF2OgJOI7iEt60mis%3D`

A full list of URLs of all training set files can be downloaded from the registration system.

Parameter	Explanation
`-i [URL file to segment]` REQUIRED	The URL of a file to segment + shared access key. The file will be in NIFTI.GZ format. The filename contains info on: PatientID, ModalityID, ModalityAbbreviation `cURL+PatientID_ModalityCounter_ModalityName.nii.gz+saKey`
`-o [output path]` REQUIRED	The full path in which output files are to be stored in: - Windows VM: temporary drive (D:) - Linux VM: /mnt/resource
`-c [ConfigurationID]` REQUIRED	The program can have up to 5 configurations, indicated by the integers 1 to 5. A configuration can be for example different sets of segmentation parameters. The participants are free to choose what constitutes a configuration.

Configuration IDs

The program should accept all five configuration IDs, and provide no output if a particular configuration ID if not implemented. During the evaluation phase, the organisers will begin with executing configuration ID 1, followed by the remaining configuration IDs in increasing order, until all configurations have been executed, or the period of weeks available for the evaluation calculations runs out. In the latter case, only the results for the configurations for which the execution completed will be provided.

Output Files

Output	Explanation
`lesion location files`	Files containing the coordinates of the detected lesions. Formating analogous to the annotation files in slicer3D convention, as specified in Section 5.2 of this document. `PatientID_ModalityCounter_ModalityName_Lesions_ParticipantID_ConfigurationID.fcsv`

Example

During the evaluation phase, the organisers will call the program of the participant with participantID **`3pp58`** as follows:

```
execute_standard -i
http://visceralstorage1.blob.core.windows.net/testset/001_1_CT_wb.nii.gz?sr=c
&si=readonly&sig=Z69O9Vz8TU0RxawtASpmpWZnT%2FhF2OgJOI7iEt60mis%3D -o
/mnt/resource/output_p1/ -c 2
```

meaning that it should produce all lesions the algorithm can find in the specified volume using configuration 2, and write the file:

```
/mnt/resource/output_p1/001_1_CT_wb_Lesions_3pp58_2.fcsv
```

5.2 Specification: lesion location file

`PatientID_ModalityCounter_ModalityName_Lesions_ParticipantID_ConfigurationID.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 will be used to calculate the evaluation metrics. As soon as the metrics to be used are implemented, an announcement will be made to participants. The latest version of the EvaluateSegmentation software is always available here: https://github.com/codalab/EvaluateSegmentation

7. FAQ

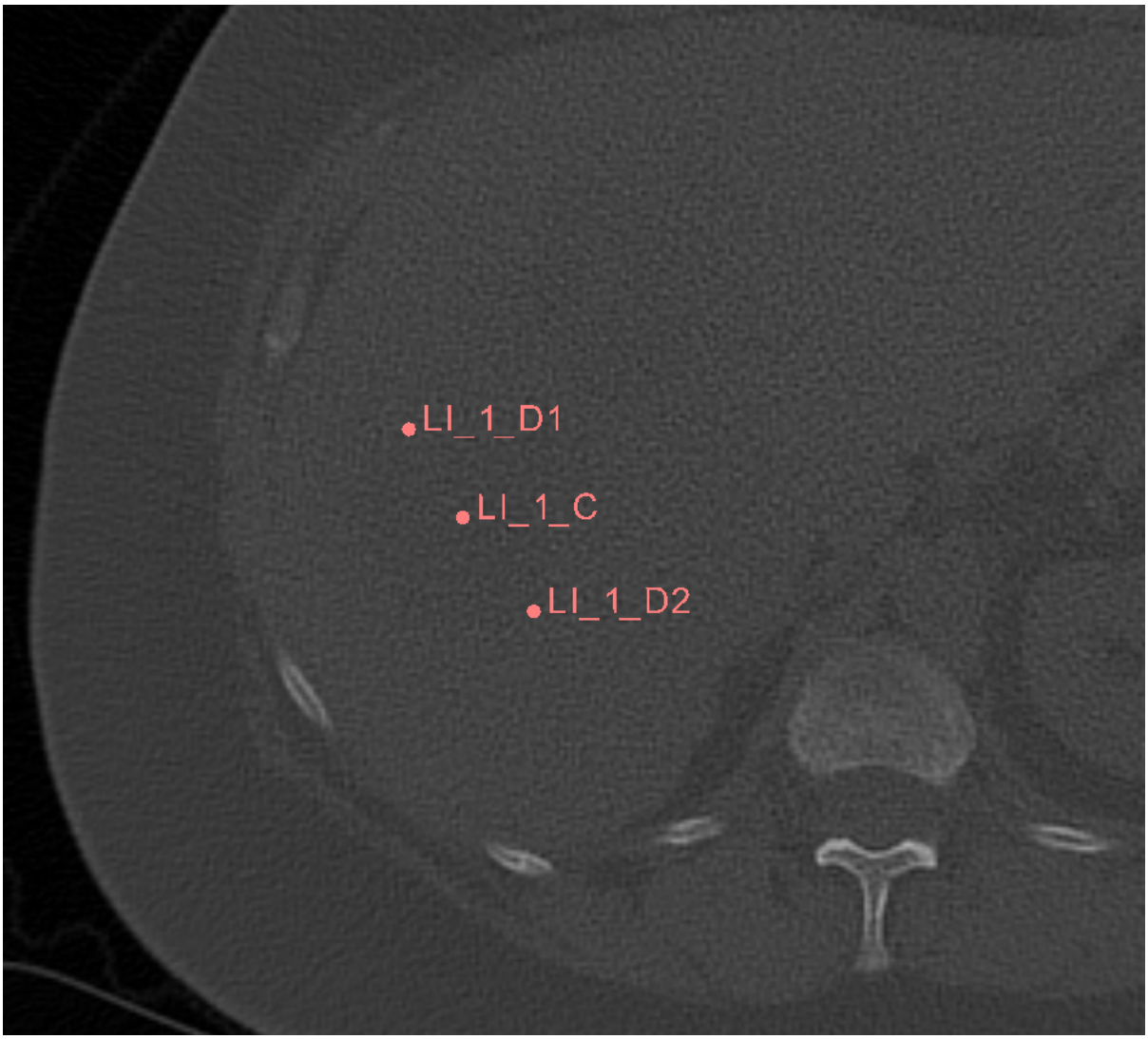
- Is it possible to participate in only part of the benchmark? i.e. detect only bone/lungs lesion.

Yes, you can participate in the benchmark only for a specific organ, if you tell us which during the submission process. We will enable this.

- 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 instructured 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.